

STATISTICAL INFERENCE FOR MISSPECIFIED ERGODIC LÉVY DRIVEN STOCHASTIC DIFFERENTIAL EQUATION MODELS

YUMA UEHARA

ABSTRACT. We consider the estimation problem of misspecified ergodic Lévy driven stochastic differential equation models based on high-frequency samples. We utilize a widely applicable and tractable Gaussian quasi-likelihood approach which focuses on mean and variance structure. It is shown that the corresponding Gaussian quasi-likelihood estimators of the drift and scale parameters satisfy polynomial type probability estimates and asymptotic normality at the same rate as the correctly specified case. In this process, the theory of extended Poisson equation for time-homogeneous Feller Markov processes plays an important role. Our result confirms the reliability of the Gaussian quasi-likelihood approach for SDE models, more firmly.

1. INTRODUCTION

Up to the present, a great deal of empirical analyses have pointed out the existence of non-Gaussian activities lying in time-varying phenomena such as the log-return of stock prices, the spike noise of neurons, and so on. To incorporate them into statistical modeling, Lévy processes, which can be interpreted as a natural continuous time version of random walks, are regarded as crucial components. Hence the statistical theory of the stochastic differential equation driven (SDE) by them (including themselves as a matter of course) based on high-frequency samples has been developed so far. Since its genuine likelihood cannot generally be obtained in a closed form, many other feasible ways have been considered, for instance, the threshold based estimation for jump diffusion models by [35] and [30], the least absolute deviation (LAD)-type estimation for Lévy driven Ornstein-Uhlenbeck models by [22], the non-Gaussian stable quasi-likelihood estimation for locally stable driven SDE models by [26], and the least square estimation for small Lévy driven SDE models by [19]. These literatures assume the precise structure of their driving noises. As for the methods not assuming it, we refer to the Gaussian quasi-likelihood (GQL) based estimation schemes proposed in [23] and [28].

However, all of the above studies premise that the data-generating model is correctly specified even though the risk of model misspecification is essentially inevitable. It should be considered in the deviation of the asymptotic properties of estimators and estimating functions such as for ensuring the reliability of the statistical methods and comparing candidate models. In this paper, we give the theoretical framework of the parametric inference for misspecified Lévy driven SDE models based on high-frequency samples for the first time. Let X be the one-dimensional stochastic process defined on the complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, P)$ such that

$$(1.1) \quad dX_t = A(X_t)dt + C(X_t)dZ_t,$$

where:

- Z is a one-dimensional càdlàg Lévy process without Wiener part. It is independent of the initial variable X_0 and satisfies $E[Z_1] = 0$, $Var[Z_1] = 1$, and $E[|Z_1|^q] < \infty$ for all $q > 0$;
- The coefficients $A : \mathbb{R} \mapsto \mathbb{R}$ and $C : \mathbb{R} \mapsto \mathbb{R}$ are Lipschitz continuous;
- $\mathcal{F}_t := \sigma(X_0) \vee \sigma(Z_s; s \leq t)$.

We suppose that the observations $(X_{t_0}, \dots, X_{t_n})$ are obtained from the solution path X in the so-called “rapidly increasing experimental design”, that is, $t_j \equiv t_j^n := jh_n$, $T_n := nh_n \rightarrow \infty$, and $nh_n^2 \rightarrow 0$. We consider the following parametric one-dimensional SDE model:

$$(1.2) \quad dX_t = a(X_t, \alpha)dt + c(X_t, \gamma)dZ_t,$$

where the functional forms of the coefficients $a : \mathbb{R} \times \Theta_\alpha \mapsto \mathbb{R}$ and $c : \mathbb{R} \times \Theta_\gamma \mapsto \mathbb{R}$ are supposed to be known except for a finite-dimensional unknown parameter $\theta := (\alpha, \gamma)$ being an element of the bounded convex domain $\Theta := \Theta_\alpha \times \Theta_\gamma \subset \mathbb{R}^p$. Here the true coefficients $(A, C)(\cdot)$ are not necessarily in the parametric

Date: April 4, 2017.

Key words and phrases. Lévy driven stochastic differential equation, misspecified model, Gaussian quasi-likelihood estimation, extended Poisson equation, high-frequency sampling, stepwise estimation.

TABLE 1. GQL approach for ergodic diffusion models and ergodic Lévy driven SDE models

Model	Rates of convergence		Ref.
	drift	scale	
correctly specified diffusion	$\sqrt{T_n}$	\sqrt{n}	[16], [37]
misspecified diffusion	$\sqrt{T_n}$	$\sqrt{T_n}$	[36]
correctly specified Lévy driven SDE	$\sqrt{T_n}$	$\sqrt{T_n}$	[24], [28]
misspecified Lévy driven SDE	$\sqrt{T_n}$	$\sqrt{T_n}$	this paper

family $\{(a, c)(\cdot, \theta) : \theta \in \Theta\}$, namely, the coefficients are possibly misspecified. Under such situation, we adopt the staged Gaussian quasi-likelihood estimation procedure used in [28] to estimate an *optimal* value θ^* of θ . The *optimality* is herein determined by minimizing the Kullback-Leibler (KL) divergence like quantities (2.1) and (2.2) below; they are the probability limits of the staged GQL. Hereinafter, the terminologies “misspecified” and “misspecification” will be used for the misspecification with respect to the coefficient unless another meaning is specifically indicated.

Concerning misspecified ergodic diffusion models, it is shown in [36] that although the misspecification with respect to their diffusion term deteriorates the convergence rate of the scale (diffusion) parameter, the Gaussian quasi-maximum likelihood estimator (GQMLE) still has asymptotic normality. In the paper, the differential equations endowed with their infinitesimal generator (cf. [31]) is the key to derive the behavior of the GQMLE. However, since the infinitesimal generator of X contains the integro-operator with respect to the Lévy measure of Z , it is difficult to verify the existence and regularity of the corresponding equation. To avoid such obstacle, we will substantially invoke the theory of the extended Poisson equation (EPE) for homogeneous Feller Markov processes established in [38]. Applying the result of [38] for (1.1), the existence and weighted Hölder regularity of the solution of EPEs will be derived under a mighty mixing condition on X . Building on the result, we will provide the asymptotic normality of our staged GQMLE and its tail probability estimates under sufficient regularity and moment conditions on the ingredients of (1.1) and (1.2). Here we note that the absence of Wiener part in (1.1) is essential while it is not in the correctly specified case, for more details, see Remark 3.8.

It will turn out that the convergence rate of the scale parameters is $\sqrt{T_n}$, and it is the same as the correctly specified case, see [23] and [28]. This is different from the diffusion case (cf. Table 1). Such difference may be caused from applying the GQL to non-Gaussian driving noises, that is, the efficiency loss of the GQMLE may occur even in the correctly specified case. Indeed, the non-Gaussian stable quasi-likelihood is known to estimate the drift and scale parameters faster than the GQMLE in correctly specified locally β -stable driven SDE models (cf. [26]); each of their convergence rates are $\sqrt{nh_n^{1-1/\beta}}$ and \sqrt{n} , respectively. Further, for correctly specified locally β -stable driven Ornstein-Uhlenbeck models, the LAD-type estimators of [22] tend to the true value at the speed of $\sqrt{nh_n^{1-1/\beta}}$ and it is also faster than that of the GQMLE. However, in exchange for its efficiency, the GQL approach is worth considering by the following reasons:

- It does not include any special functions (e.g. Bessel function, Whittaker function, and so on), infinite expansion series and analytically unsolvable integrals, thus computation based on it is not relatively time-consuming.
- It focuses only on the (conditional) mean and covariance structure, thus it does not need so much restriction on the driving noise and is robust against the noise misspecification. In other words, we can construct reasonable estimators of the drift and scale coefficients in the unified way if only the driving Lévy noise has moments of any order.

Our result ensures that even if the coefficients are misspecified, the GQL still works for Lévy driven SDE models and completely inherits its merit written in above.

The rest of this paper is organized as follows: In Section 2, we introduce our estimation procedure and assumptions. Section 3 provides our main results in the following turn:

- (1) the tail probability estimates of the GQMLE;
- (2) the existence and weighted Hölder regularity of the solution of EPEs for Lévy driven SDEs;
- (3) the asymptotic normality of the GQMLE at $\sqrt{T_n}$ -rate.

A simple numerical experiment is presented in Section 4. We give all proofs of our results in Section 5.

2. ESTIMATION SCHEME AND ASSUMPTIONS

For notational convenience, we will use the following symbols without any mention:

- ∂_x is referred to as a differential operator for any variable x .
- $x_n \lesssim y_n$ implies that there exists a positive constant C being independent of n satisfying $x_n \leq Cy_n$ for all large enough n .
- \bar{S} denotes the closure of any set S .
- We write $Y_j = Y_{t_j}$ and $\Delta_j Y = Y_j - Y_{j-1}$ for any stochastic process Y .
- For any matrix valued function f on $\mathbb{R} \times \Theta$, we write $f_s(\theta) = f(X_s, \theta)$; especially we write $f_j(\theta) = f(X_j, \theta)$.
- η stands for the law of X_0 .
- $E^j[\cdot]$ denotes the conditional expectation with respect to \mathcal{F}_{t_j} .

We define our staged GQMLE $\hat{\theta}_n := (\hat{\alpha}_n, \hat{\gamma}_n)$ in the following manner:

- (1) *Drift-free estimation of γ* . Define the Maximizing-type estimator (so-called M -estimator) $\hat{\gamma}_n$ by

$$\hat{\gamma}_n \in \operatorname{argmax}_{\gamma \in \Theta_\gamma} \mathbb{G}_{1,n}(\gamma),$$

for the \mathbb{R} -valued random function

$$\mathbb{G}_{1,n}(\gamma) := -\frac{1}{T_n} \sum_{j=1}^n \left\{ h_n \log c_{j-1}^2(\gamma) + \frac{(\Delta_j X)^2}{c_{j-1}^2(\gamma)} \right\}.$$

- (2) *Weighted least square estimation of α* . Define the least square type estimator $\hat{\alpha}_n$ by

$$\hat{\alpha}_n \in \operatorname{argmax}_{\alpha \in \Theta_\alpha} \mathbb{G}_{2,n}(\alpha),$$

for the \mathbb{R} -valued random function

$$\mathbb{G}_{2,n}(\alpha) := -\frac{1}{T_n} \sum_{j=1}^n \frac{(\Delta_j X - h_n a_{j-1}(\alpha))^2}{h_n c_{j-1}^2(\hat{\gamma}_n)}.$$

Let $\mathbb{G}_1 : \Theta_\gamma \mapsto \mathbb{R}$ and $\mathbb{G}_2 : \Theta_\alpha \mapsto \mathbb{R}$ be

$$(2.1) \quad \mathbb{G}_1(\gamma) := - \int_{\mathbb{R}} \left(\log c^2(x, \gamma) + \frac{C^2(x)}{c^2(x, \gamma)} \right) \pi_0(dx),$$

$$(2.2) \quad \mathbb{G}_2(\alpha) := - \int_{\mathbb{R}} c(x, \gamma^*)^{-2} (A(x) - a(x, \alpha))^2 \pi_0(dx),$$

where π_0 denotes the probability measure on \mathbb{R} introduced later. For these functions, we define an *optimal* value $\theta^* := (\alpha^*, \gamma^*)$ of θ by

$$\gamma^* \in \operatorname{argmax}_{\gamma \in \Theta_\gamma} \mathbb{G}(\gamma), \quad \alpha^* \in \operatorname{argmax}_{\alpha \in \Theta_\alpha} \mathbb{G}(\alpha).$$

We assume that θ^* is unique, and that it is in the interior of Θ .

Remark 2.1. *Although our estimation method ignores the drift term in the first stage, the effect of it asymptotically vanishes. This is because the scale term dominates the small time behavior of X in L_2 -sense. Specifically, we can derive*

$$E^{j-1} \left[\left(\int_{t_{j-1}}^{t_j} f_{s-}(\theta) dJ_s \right)^2 \right] \lesssim h_n f_{j-1}^2, \quad E^{j-1} \left[\left(\int_{t_{j-1}}^{t_j} g_s(\theta) ds \right)^2 \right] \lesssim h_n^2 g_{j-1}^2,$$

for suitable functions f and g . Indeed, it has already been shown that the asymptotic behavior of the scale estimator constructed by our manner is the same as the conventional GQL estimator in the case of correctly specified ergodic diffusion models (cf. [37]) and ergodic Lévy driven SDE models (cf. [28]). Such ignorance should be helpful in reducing the number of simultaneous optimization parameters, thus our estimators are expected to numerically be more stabilized and their calculation should be less time-consuming. Moreover, by choosing appropriate functional forms, each estimation stage is reduced to a convex optimization problem. For example, if a and c are linear and log-linear with respect to parameters, respectively, then the above argument holds. As for other candidates of (a, c) and details, see [28, Example 3.8].

Remark 2.2. From the functional form of \mathbb{G}_1 and \mathbb{G}_2 , the maximizers of them correspond with the true parameters in the correctly specified case. Hence an intuitive interpretation of θ^* is the parameter value which yields the closest model to the data-generating model measured by the KL divergence like quantities (2.1) and (2.2). The estimation procedure at the first stage can be regarded as Stein's loss function (or Jensen-Bregman LogDet divergence) based estimation, too. Thus θ^* can also be interpreted as the minimizer of each expected loss (Stein's loss and a weighted L_2 loss) with respect to π_0 .

To derive our asymptotic results, we introduce some assumptions below. Let ν_0 be the Lévy measure of Z .

Assumption 2.3. (1) $E[Z_1] = 0$, $\text{Var}[Z_1] = 1$, and $E[|Z_1|^q] < \infty$ for all $q > 0$.
 (2) The Blumenthal-Gettoor index (BG-index) of Z is smaller than 2, that is,

$$\beta := \inf \left\{ \gamma \geq 0 : \int_{|z| \leq 1} |z|^\gamma \nu_0(dz) < \infty \right\} < 2.$$

We remark that the condition on the BG-index does not excessively restrict the candidates of Z , let alone the first condition. Indeed, there are a lot of Lévy processes satisfying Assumption 2.3, for example, bilateral gamma process, normal tempered stable process, normal inverse Gaussian process, variance gamma process, and so on.

Assumption 2.4. (1) The coefficients $A(\cdot)$ and $C(\cdot)$ are Lipschitz continuous and twice differentiable, and their first and second derivatives are of at most polynomial growth.
 (2) The drift coefficient $a(\cdot, \alpha^*)$ and scale coefficient $c(\cdot, \gamma^*)$ are Lipschitz continuous, and $c(\cdot, \cdot)$ is invertible for all (x, γ) .
 (3) For each $i \in \{0, \dots, 1\}$ and $k \in \{0, \dots, 5\}$, the following conditions hold:
 • The coefficients a and c admit extension in $\mathcal{C}(\mathbb{R} \times \bar{\Theta})$ and have the partial derivatives $(\partial_x^i \partial_\alpha^k a, \partial_x^i \partial_\gamma^k c)$ possessing extension in $\mathcal{C}(\mathbb{R} \times \bar{\Theta})$.
 • There exists nonnegative constant $C_{(i,k)}$ satisfying

$$(2.3) \quad \sup_{(x, \alpha, \gamma) \in \mathbb{R} \times \bar{\Theta}_\alpha \times \bar{\Theta}_\gamma} \frac{1}{1 + |x|^{C_{(i,k)}}} \{ |\partial_x^i \partial_\alpha^k a(x, \alpha, \gamma)| + |\partial_x^i \partial_\gamma^k c(x, \gamma)| + |c^{-1}(x, \gamma)| \} < \infty.$$

We note that the first part of Assumption 2.3 and Assumption 2.4 ensures the existence of the strong solution of SDE (1.1), that is, there exists a measurable function g such that $X = g(X_0, Z)$.

Given a function $\rho : \mathbb{R} \rightarrow \mathbb{R}^+$ and a signed measure m on a one-dimensional Borel space, we define

$$\|m\|_\rho = \sup \{ |m(f)| : f \text{ is } \mathbb{R}\text{-valued, } m\text{-measurable and satisfies } |f| \leq \rho \}.$$

Hereinafter $P_t(x, \cdot)$ denotes the transition probability of X .

Assumption 2.5. (1) There exists a probability measure π_0 such that for every $q > 0$ we can find constants $a > 0$ and $C_q > 0$ for which

$$(2.4) \quad \sup_{t \in \mathbb{R}_+} \exp(at) \|P_t(x, \cdot) - \pi_0(\cdot)\|_{h_q} \leq C_q h_q(x), \quad x \in \mathbb{R},$$

$$\text{where } h_q(x) := 1 + |x|^q.$$

(2) For all $q > 0$, we have

$$(2.5) \quad \sup_{t \in \mathbb{R}_+} E_0[|X_t|^q] < \infty.$$

Assumption 2.6. There exist positive constants χ_γ and χ_α such that for all $(\gamma, \alpha) \in \Theta$,

$$\begin{aligned} \mathbb{Y}_1(\gamma) &:= \mathbb{G}_1(\gamma) - \mathbb{G}_1(\gamma^*) \leq -\chi_\gamma |\gamma - \gamma^*|^2, \\ \mathbb{Y}_2(\alpha) &:= \mathbb{G}_2(\alpha) - \mathbb{G}_2(\alpha^*) \leq -\chi_\alpha |\alpha - \alpha^*|^2. \end{aligned}$$

We introduce a $p \times p$ -matrix $\Gamma := \begin{pmatrix} \Gamma_\gamma & O \\ \Gamma_{\alpha\gamma} & \Gamma_\alpha \end{pmatrix}$ whose components are defined by:

$$\begin{aligned} \Gamma_\gamma &:= 2 \int_{\mathbb{R}} \frac{\partial_\gamma^{\otimes 2} c(x, \gamma^*) c(x, \gamma^*) - (\partial_\gamma c(x, \gamma^*))^{\otimes 2}}{c^4(x, \gamma^*)} (C^2(x) - c^2(x, \gamma^*)) \pi_0(dx) \\ &\quad - 4 \int_{\mathbb{R}} \frac{(\partial_\gamma c(x, \gamma^*))^{\otimes 2}}{c^4(x, \gamma^*)} C^2(x) \pi_0(dx), \\ \Gamma_{\alpha\gamma} &:= 2 \int_{\mathbb{R}} \partial_\alpha a(x, \alpha^*) \partial_\gamma^\top c^{-2}(x, \gamma^*) (A(x) - a(x, \alpha^*)) \pi_0(dx), \end{aligned}$$

$$\Gamma_\alpha := 2 \int_{\mathbb{R}} \frac{\partial_\alpha^{\otimes 2} a(x, \alpha^*)}{c^2(x, \gamma^*)} (a(x, \alpha^*) - A(x)) \pi_0(dx) + 2 \int_{\mathbb{R}} \frac{(\partial_\alpha a(x, \alpha^*))^{\otimes 2}}{c^2(x, \gamma^*)} \pi_0(dx),$$

where $x^{\otimes 2} := x^\top x$ for any vector x .

Assumption 2.7. Γ is invertible.

3. MAIN RESULTS

In this section, we state our main results only for the fully misspecified case, that is, both of the true coefficients $A(\cdot)$ and $C(\cdot)$ do not belong to the parametric family $\{(a, c)(\cdot, \theta) : \theta \in \Theta\}$. Concerning the partly misspecified case (i.e. for either of A and C is specified), similar results can be derived just as the corollaries (see, Remark 3.7). All of their proofs will be given in Appendix. The first result provides the tail probability estimates of the normalized $\hat{\theta}_n$:

Theorem 3.1. *Suppose that Assumptions 2.3-2.7 hold. Then, for any $L > 0$ and $r > 0$, there exists a positive constant C_L such that*

$$(3.1) \quad \sup_{n \in \mathbb{N}} P(|\sqrt{T_n}(\hat{\theta}_n - \theta^*)| > r) \leq \frac{C_L}{r^L}.$$

Polynomial-type tail probability estimates ensure the moment convergence of the corresponding estimators which is theoretically essential such as in the deviation of an information criterion, residual analysis, and the measurement of L_q -prediction error. To establish Theorem 3.1, we will rely on the theory of *polynomial-type large deviation inequality* (PLDI) introduced by [42]. In order to achieve the sufficient condition for the PLDI, the moment bounds and strong identifiability conditions with respect to the corresponding estimating functions are needed. These conditions mostly imply the asymptotic normality of the corresponding M -estimators based on Markov-type estimating functions $M_n(\theta) = \sum_{j=1}^n m(X_{j-1}, X_j; \theta)$. However, such conditions are insufficient for our case. More specifically, we will face the following situation:

((scaled) quasi-score function) = (sum of martingale difference) + (intractable term) + (negligible term),
where the intractable term is expressed as

$$\sqrt{\frac{h_n}{n}} \sum_{j=1}^n f_{j-1}(\theta^*) = \frac{1}{\sqrt{T_n}} \int_0^{T_n} f_s(\theta^*) ds + o_p(1),$$

with a specific measurable function f satisfying $\pi_0(f) = 0$. The celebrated CLT-type theorems for such single integral functional form of Markov processes have been reported in many literatures, for example, [6, Theorem 2.1], [15, Theorem VIII 3.65], [17, Theorem 2.1], [39, Corollary 4.1], and the references therein. However the combination with the original leading term makes it difficult to clarify the asymptotic behavior of the left-hand-side. To handle this difficulty, we invoke the concept of the EPE introduced in [38]:

Definition 3.2. [38, Definition 2.1] *We say that a measurable function $f : \mathbb{R} \rightarrow \mathbb{R}$ belongs to the domain of the extended generator \tilde{A} of a càdlàg homogeneous Feller Markov process Y taking values in \mathbb{R} if there exists a measurable function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that the process*

$$f(Y_t) - \int_0^t g(Y_s) ds, \quad t \in \mathbb{R}^+,$$

is well defined and is a local martingale with respect to the natural filtration of Y and every measure $P_x(\cdot) := P(\cdot | Y_0 = x)$, $x \in \mathbb{R}$. For such a pair (f, g) , we write $f \in \text{Dom}(\tilde{A})$ and $\tilde{A}f \stackrel{\text{EPE}}{=} g$.

Remark 3.3. *In the previous definition, the terminology “Feller” means that the corresponding transition semigroup T_t is a mapping $C_b(\mathbb{R})$ into $C(\mathbb{R})$. When it comes to X , its homogeneous, Feller and (strong) Markov properties are guaranteed by the argument in [2, Theorem 6.4.6] and [21, 3.1.1 (ii)].*

Hereinafter $y^{(i)}$ is referred to as the i -th component of any vector y . We consider the following EPEs:

$$(3.2) \quad \tilde{A}u_1^{(j_1)}(x) \stackrel{\text{EPE}}{=} -\frac{\partial_{\gamma^{(j_1)}} c(x, \gamma^*)}{c^3(x, \gamma^*)} (c^2(x, \gamma^*) - C^2(x)),$$

$$(3.3) \quad \tilde{A}u_2^{(j_2)}(x) \stackrel{\text{EPE}}{=} -\frac{\partial_{\alpha^{(j_2)}} a(x, \alpha^*)}{c^2(x, \gamma^*)} (A(x) - a(x, \alpha^*)),$$

for the extended generator $\tilde{\mathcal{A}}$ of X , $j_1 \in \{1, \dots, p_\gamma\}$ and $j_2 \in \{1, \dots, p_\alpha\}$. Henceforth E^x is referred to as the expectation operator with the initial condition $X_0 = x$, that is,

$$E^x[g(X_t)] = \int_{\mathbb{R}} g(y) P_t(x, dy),$$

for any measurable function g . The next proposition ensures the existence of the solutions and verifies their behavior:

Proposition 3.4. *Under Assumption 2.3-2.5, there exist unique solutions of (3.2) and (3.3), and the solution vectors $f_1 := (f_1^{(j_1)})_{j_1 \in \{1, \dots, p_\gamma\}}$ and $f_2 := (f_2^{(j_2)})_{j_2 \in \{1, \dots, p_\alpha\}}$ satisfy*

$$\sup_{x, y \in \mathbb{R}} \frac{|f_i(x) - f_i(y)|}{|x - y|^{1/p_i} (1 + |x|^{q_i L_i} + |y|^{q_i L_i})} < \infty, \quad \text{for } i \in \{1, 2\},$$

where any $p_i \in (1, \infty)$, $q_i = p_i/(p_i - 1)$, and some positive constants L_1 and L_2 . Furthermore, $f_1(X_t) + \int_0^t \partial_\gamma c(X_s, \gamma^*) (c^2(X_s, \gamma^*) - C^2(X_s)) / c^3(X_s, \gamma^*) ds$ and $f_2(X_t) + \int_0^t \partial_\alpha a(X_s, \alpha^*) (A(X_s) - a(X_s, \alpha^*)) / c^2(X_s, \gamma^*) ds$ are L_2 -martingale with respect to (\mathcal{F}_t, P_x) for every $x \in \mathbb{R}$, and their explicit forms are given as follows:

$$\begin{aligned} f_1(x) &= \int_0^\infty E^x \left[\frac{\partial_\gamma c(X_t, \gamma^*)}{c^3(X_t, \gamma^*)} (c^2(X_t, \gamma^*) - C^2(X_t)) \right] dt, \\ f_2(x) &= \int_0^\infty E^x \left[\frac{\partial_\alpha a(X_t, \alpha^*)}{c^2(X_t, \gamma^*)} (A(X_t) - a(X_t, \alpha^*)) \right] dt. \end{aligned}$$

Remark 3.5. *Thanks to the result of the previous theorem and assumptions on the coefficients, $f_1(X_t) + \int_0^t \partial_\gamma c(X_s, \gamma^*) (c^2(X_s, \gamma^*) - C^2(X_s)) / c^3(X_s, \gamma^*) ds$ and $f_2(X_t) + \int_0^t \partial_\alpha a(X_s, \alpha^*) (A(X_s) - a(X_s, \alpha^*)) / c^2(X_s, \gamma^*) ds$ have finite second-order moments. Thus, slightly refining the argument in [33, the proof of Proposition VI I 1.6] with the monotone convergence theorem, the L_2 -martingale property of them with respect to (\mathcal{F}_t, P_x) can be replaced by the L_2 -martingale property with respect to (\mathcal{F}_t, P) in the previous proposition.*

Building on the previous proposition, now we can obtain the asymptotic normality of $\sqrt{T_n}(\hat{\theta}_n - \theta^*)$:

Theorem 3.6. *Under Assumptions 2.3-2.7, there exists a nonnegative definite matrix $\Sigma \in \mathbb{R}^p \otimes \mathbb{R}^p$ such that*

$$\sqrt{T_n}(\hat{\theta}_n - \theta^*) \xrightarrow{\mathcal{L}} N(0, \Gamma^{-1} \Sigma (\Gamma^{-1})^\top),$$

and the form of $\Sigma := \begin{pmatrix} \Sigma_\gamma & \Sigma_{\alpha\gamma} \\ \Sigma_{\alpha\gamma}^\top & \Sigma_\alpha \end{pmatrix}$ is given by:

$$\begin{aligned} \Sigma_\gamma &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma^*)}{c^3(x, \gamma^*)} C^2(x) z^2 + f_1(x + C(x)z) - f_1(x) \right)^{\otimes 2} \pi_0(dx) \nu_0(dz), \\ \Sigma_{\alpha\gamma} &= -4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma^*)}{c^3(x, \gamma^*)} C^2(x) z^2 + f_1(x + C(x)z) - f_1(x) \right) \\ &\quad \left(\frac{\partial_\alpha a(x, \alpha^*)}{c^2(x, \gamma^*)} C(x)z + f_2(x + C(x)z) - f_2(x) \right)^\top \pi_0(dx) \nu_0(dz), \\ \Sigma_\alpha &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\alpha a(x, \alpha^*)}{c^2(x, \gamma^*)} C(x)z + f_2(x + C(x)z) - f_2(x) \right)^{\otimes 2} \pi_0(dx) \nu_0(dz). \end{aligned}$$

Remark 3.7. *If either of the coefficients is correctly specified, the right-hand side of the associated EPE (3.2) or (3.3) is identically 0. Thus we have*

$$\begin{aligned} \Sigma_\gamma &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma_0)}{c(x, \gamma_0)} \right)^{\otimes 2} \pi_0(dx) \int_{\mathbb{R}} z^4 \nu_0(dz), \\ \Sigma_{\alpha\gamma} &= -4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma_0)}{c(x, \gamma_0)} z^2 \right) \left(\frac{\partial_\alpha a(x, \alpha^*)}{c(x, \gamma_0)} z + f_2(x + c(x, \gamma_0)z) - f_2(x) \right)^\top \pi_0(dx) \nu_0(dz), \\ \Sigma_\alpha &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\alpha a(x, \alpha^*)}{c(x, \gamma_0)} z + f_2(x + c(x, \gamma_0)z) - f_2(x) \right)^{\otimes 2} \pi_0(dx) \nu_0(dz), \end{aligned}$$

in the case that the scale coefficient is correctly specified and

$$\begin{aligned} \Sigma_\gamma &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma^*)}{c^3(x, \gamma^*)} C^2(x) z^2 + f_1(x + C(x)z) - f_1(x) \right)^{\otimes 2} \pi_0(dx) \nu_0(dz), \\ \Sigma_{\alpha\gamma} &= -4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma^*)}{c^3(x, \gamma^*)} C^2(x) z^2 + f_1(x + C(x)z) - f_1(x) \right) \left(\frac{\partial_\alpha a(x, \alpha_0)}{c^2(x, \gamma^*)} C(x)z \right)^\top \pi_0(dx) \nu_0(dz), \end{aligned}$$

$$\Sigma_\alpha = 4 \int_{\mathbb{R}} \left(\frac{\partial_\alpha a(x, \alpha_0)}{c^2(x, \gamma^*)} C(x) \right)^{\otimes 2} \pi_0(dx),$$

in the case that the drift coefficient is correctly specified. In above equations, γ_0 and α_0 are referred to as the true values, namely, they satisfy that $c(x, \gamma_0) \equiv C(x)$ and $a(x, \alpha_0) \equiv A(x)$ on the whole state space of X .

Remark 3.8. In this remark, we suppose that the data-generating model defined on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, P)$ is supposed to be

$$(3.4) \quad dY_t = A(Y_t)dt + B(Y_t)dW_t + C(Y_{t-})dZ_t,$$

where W is a standard Wiener process independent of (Y_0, Z) , $\mathcal{F}_t := \sigma(Y_0) \vee \sigma((W_s, Z_s); s \leq t)$ and $B : \mathbb{R} \rightarrow \mathbb{R}$ is a measurable function. We look at the following parametric model:

$$dY_t = a(Y_t, \alpha)dt + b(Y_t, \gamma)dW_t + c(Y_{t-}, \gamma)dZ_t,$$

where $b : \mathbb{R} \times \Theta_\gamma \rightarrow \mathbb{R}$ is a measurable function. Here other ingredients are similarly defined as above and we use the same notations for its transition probability, invariant measure, and so on. When the true coefficients (A, B, C) are correctly specified, the GQMLE still has asymptotic normality and the sufficient conditions for it are easy to check (cf. [23]). However, we note that it is difficult to give such conditions when they are misspecified. This is because our methodology based on the EPE becomes insufficient due to the presence of Wiener component in the deviation of the asymptotic variance (see, the proof of Theorem 3.6). To formally derive a similar result to Theorem 3.6, we may additionally have to impose the following condition:

Condition A: There exists a unique C^2 -solution f on \mathbb{R} of

$$(3.5) \quad Af(x) = A(x)\partial_x f(x) - \frac{1}{2}B(x)\partial_x^2 f(x) + \int_{\mathbb{R}} (f(x+C(x)z) - f(x) - \partial_x f(x)C(x)z)\nu_0(dz) = g(A(x), B(x), C(x)),$$

where $g(A(x), B(x), C(x))$ is a specific function satisfying $\int_{\mathbb{R}} g(A(x), B(x), C(x))\pi_0(dx) = 0$. Furthermore, the first and second derivatives of f are of at most polynomial growth.

Under **Condition A**, the limit distribution of the GQMLE can be derived by combining the proof of [37] and Theorem 3.6. It is known that the theory of viscosity solutions for integro-differential equations ensures the existence of f in limited situation, for instance, see [3], [4], [13] and [14]. However, it is not so for the regularity of f . As another attempt to confirm **Condition A**, the associated EPE $\tilde{A}\tilde{f} \stackrel{EPE}{=} g$ may possibly be helpful. This is because the existence and uniqueness of the solution \tilde{f} of the EPE can be verified in an analogous way to Theorem 3.4, and if \tilde{f} admits C^2 -property and growth conditions in **Condition A**, then \tilde{f} satisfies (3.5). The latter argument can formally be shown as follows:

It is enough to check $A\tilde{f} = g$. Since $\tilde{f}(Y_t) - \int_0^t g(A(Y_s), B(Y_s), C(Y_s))ds$ is a martingale with respect to (\mathcal{F}_t, P_x) for all $x \in \mathbb{R}$, we have

$$E^x \left[\tilde{f}(Y_t) - \int_0^t g(A(Y_s), B(Y_s), C(Y_s))ds \right] = \tilde{f}(x).$$

Hence it follows from Itô's formula that as $t \rightarrow 0$,

$$\begin{aligned} \left| \frac{E^x[\tilde{f}(Y_t)] - \tilde{f}(x)}{t} - g(A(x), B(x), C(x)) \right| &= \left| \frac{1}{t} \int_0^t (E^x[g(A(Y_s), B(Y_s), C(Y_s))] - g(A(x), B(x), C(x))) ds \right| \\ &= \left| \frac{1}{t} \int_0^t \int_0^s E^x[A g(A(Y_u), B(Y_u), C(Y_u))]duds \right| \lesssim t \rightarrow 0. \end{aligned}$$

In this sketch, we implicitly assume suitable regularity and moment conditions on each ingredient, but they are reduced to be conditions on the true coefficients (A, B, C) . Thus, verifying the behavior of

$$\tilde{f}(x) = \int_0^\infty E^x[g(A(Y_t), B(Y_t), C(Y_t))]dt = \int_0^\infty \int_{\mathbb{R}} g(A(y), B(y), C(y))P_t(x, dy)$$

leads to **Condition A**. Just for Lévy driven Ornstein-Uhlenbeck models, we can observe the property of $\tilde{f}(x) = \int_0^\infty E^x[g(A(Y_t), B(Y_t), C(Y_t))]dt$ based on the explicit form of the solution (cf. Example 3.9). Although, for general Lévy driven SDEs, the gradient estimates of their transition probability making use of Malliavin calculus have been investigated lately (cf. [40], [41], and the references therein), the property of $\tilde{f}(x) = \int_0^\infty E^x[g(A(Y_t), B(Y_t), C(Y_t))]dt$ is still difficult to be checked as far as the author knows. Since these are out of range of this paper, we will not treat them later.

Example 3.9. Here we consider the following Ornstein-Uhlenbeck model:

$$dX_t = -\alpha X_t dt + dZ_t,$$

for a Lévy process Z not necessarily being pure-jump type and a positive constant α . Applying Itô's formula to $\exp(\alpha t)X_t$, we have $X_t = X_0 \exp(-\alpha t) + \int_0^t \exp(\alpha(s-t))dZ_s$ and $E^x[f(X_t)] = \int_{\mathbb{R}} f(x \exp(-\alpha t) + y)p_t(dy)$ for a suitable function f . Here p_t is the probability distribution function of $\int_0^t \exp(\alpha(s-t))dZ_s$ whose characteristic function \hat{p}_t is given by:

$$(3.6) \quad \hat{p}_t(u) = \exp \left\{ \int_0^t \psi(\exp(\alpha(s-t))u) ds \right\},$$

for $\psi(u) := \log E[\exp(iuZ_1)]$ (cf. [34, Theorem 3.1]). In this case, X fulfills Assumption 2.5 provided that Assumption 2.3-(1) holds, and that the Lévy measure ν_0 of Z has a continuously differentiable positive density g on an open neighborhood around the origin (for more details, see [23, Section 5]). Under such condition, if f is differentiable and itself and its derivative are of at most polynomial growth, we have

$$\begin{aligned} \left| \partial_x \left(\int_0^\infty E^x[f(X_t)] dt \right) \right| &= \left| \int_0^\infty \left(\int_{\mathbb{R}} \partial_x f(x \exp(-\alpha t) + y) p_t(dy) \right) \exp(-\alpha t) dt \right| \\ &\lesssim \int_0^\infty \{1 + |x|^K + (1 + |x|^{2K}) \exp(-\alpha t)\} \exp(-\alpha t) dt \lesssim 1 + |x|^{2K}, \end{aligned}$$

for a positive constant K . We can derive similar estimates with respect to its higher-order derivatives in the same way.

Let J be a Lévy process such that its moments of any-order exists and its triplet is $(0, b, \nu^J)$ (cf. [2]). Here b is allowed to be 0. Mimicking the previous example, we write p_t^J as the probability distribution function of $\int_0^t \exp(\alpha(s-t))dJ_s$ for a positive constant $\alpha > 0$ and $\psi^J(u)$ stands for $\log E[\exp(iuJ_1)]$ below. Combining the argument in Remark 3.8 and Example 3.9, we obtain the following corollary:

Corollary 3.10. For a natural number $k \geq 2$, let f be a polynomial growth C^k -function whose derivatives are of at most polynomial growth. Suppose that the integral of f with respect to the Borel probability measure π_0 whose characteristic function is $\exp \left\{ \int_0^\infty \psi^J(\exp(-\alpha s)u) ds \right\}$ is 0, and that ν^J has a continuously differentiable positive density on an open neighborhood around the origin. Then, the function $g(x) := \int_0^\infty E^x[f(x \exp(-\alpha t) + \int_0^t \exp(\alpha(s-t))dJ_s)] dt = \int_0^\infty \int_{\mathbb{R}} f(x \exp(-\alpha t) + y) p_t^J(dy) dt$ on \mathbb{R} is the unique solution of the following (first or second order) integro-differential equation

$$(3.7) \quad -\alpha x \partial_x g(x) - \frac{1}{2} b \partial_x^2 g(x) + \int_{\mathbb{R}} (g(x+z) - g(x) - \partial g(x)z) \nu^J(dz) = f(x),$$

and moreover, g is also a polynomial growth C^k -function.

Remark 3.11. If the Lévy measure ν^J is symmetric (i.e. the imaginary part of ψ^J is 0), the equation (3.7) is solvable for many odd functions f as a matter of course. More specifically, for $k \in \mathbb{N}$ and $f(x) = x^{2k+1}$, the solution g is

$$\begin{aligned} g(x) &= \int_0^\infty \int_{\mathbb{R}} (x \exp(-\alpha t) + y)^{2k+1} p_t^J(dy) dt \\ &= \int_0^\infty \int_{\mathbb{R}} \sum_{i=0}^{2k+1} \binom{2k+1}{i} C_i (x \exp(-\alpha t))^i y^{2k+1-i} p_t^J(dy) dt. \end{aligned}$$

By observing the derivatives of the characteristic function, $\int_{\mathbb{R}} y^{2k+1-i} p_t^J(dy)$ can be expressed by the moments of J , hence the explicit expression of g is available.

Remark 3.12. Beside the estimation of θ , what is of special interest is the inference for ν_0 which may often be an infinite dimensional parameter. Even for (A, C) being constant and specified (i.e. X is a Lévy process with drift), it may be interest in its own right and enormous papers have addressed this problem so far. We refer to [25] for comprehensive accounts under Z being assumed to have a certain parametric structure. As for the situation where just a few information on Z is available, one of plausible attempts is the method of moments proposed in [11], [12], and [29], for example. Especially [29] established a Donsker-type functional limit theorem for empirical processes arising from high-frequently observed Lévy processes. When the coefficients A and C are nonlinear functions but specified, the residual based method of moments for ν_0 by [27] is effective: using the GQMLE $\hat{\theta}_n := (\hat{\alpha}_n, \hat{\gamma}_n)$, we have

$$\frac{1}{T_n} \sum_{j=1}^n \varphi \left(\frac{\Delta_j X - h_n a_{j-1}(\hat{\alpha}_n)}{c_{j-1}(\hat{\gamma}_n)} \right) \xrightarrow{p} \int_{\mathbb{R}} \varphi(z) \nu_0(dz),$$

FIGURE 1. The plot of the density functions of (i) $NIG(10, 0, 10, 0)$ (black dotted line), (ii) $bGamma(1, \sqrt{2}, 1, \sqrt{2})$ (green line), (iii) $NIG(25/3, 20/3, 9/5, -12/5)$ (blue line), and $N(0, 1)$ (red line).

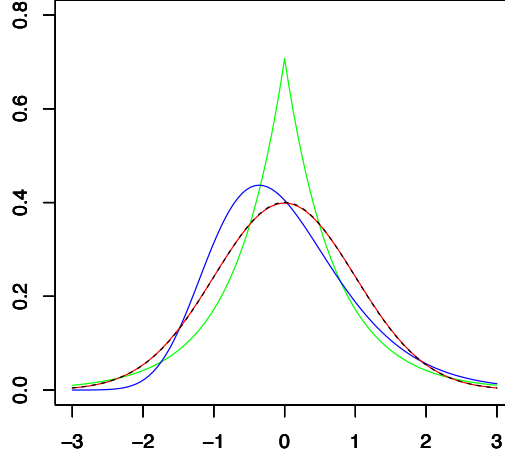


TABLE 2. The performance of our estimators; the mean is given with the standard deviation in parenthesis. The target optimal values are given in the first line of each items.

T_n	n	h_n	(i) (0.33, 1.41)		(ii) (0.37, 1.41)		(iii) (0.37, 1.41)		diffusion (0.33, 1.41)	
			$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$
50	1000	0.05	0.38	1.41	0.40	1.39	0.40	1.39	0.38	1.41
			(0.12)	(0.11)	(0.16)	(0.29)	(0.15)	(0.19)	(0.13)	(0.10)
100	5000	0.02	0.37	1.41	0.39	1.39	0.38	1.39	0.36	1.41
			(0.09)	(0.08)	(0.11)	(0.23)	(0.11)	(0.15)	(0.09)	(0.08)
100	10000	0.01	0.36	1.41	0.37	1.39	0.38	1.40	0.36	1.41
			(0.08)	(0.07)	(0.09)	(0.22)	(0.10)	(0.15)	(0.08)	(0.07)

$$\hat{D}_n \sqrt{T_n} \left(\frac{\hat{\theta}_n - \theta_0}{\frac{1}{T_n} \sum_{j=1}^n \varphi \left(\frac{\Delta_j X - h_n a_{j-1}(\hat{\alpha}_n)}{c_{j-1}(\hat{\gamma}_n)} \right)} - \int \varphi(z) \nu_0(dz) \right) \xrightarrow{\mathcal{L}} N(0, I_{p+q}),$$

for an appropriate \mathbb{R}^q -valued function φ and a $(p+q) \times (p+q)$ matrix \hat{D}_n which can be constructed only by the observations. For instance, we can choose $\varphi(z) = z^r$ and $\varphi(z) = \exp(iuz) - 1 - iuz$ (to estimate the r -th cumulant of Z and the cumulant function of Z , respectively) as φ ; see [27, Assumption 2.7] for the precise conditions on φ . As for misspecified case, if the misspecification is confined within the drift coefficient, then this scheme is still valid thanks to the faster diminishment of the mean activity in small time (cf. Remark 2.1).

4. NUMERICAL EXPERIMENTS

We suppose that the data-generating model is the following Lévy driven Ornstein-Uhlenbeck model:

$$dX_t = -\frac{1}{2}X_t dt + dZ_t, \quad X_0 = 0,$$

and that the parametric model is described as:

$$dX_t = \alpha(1 - X_t)dt + \frac{\gamma}{\sqrt{1 + X_t^2}} dZ_t.$$

The functional form of the coefficients is the same in [36, Example 3.1]. We conduct numerical experiments in three situations: (i) $\mathcal{L}(Z_t) = NIG(10, 0, 10t, 0)$, (ii) $\mathcal{L}(Z_t) = bGamma(t, \sqrt{2}, t, \sqrt{2})$, and (iii) $\mathcal{L}(Z_t) = NIG(25/3, 20/3, 9/5t, -12/5t)$. NIG (normal inverse Gaussian) random variable is defined by the normal

FIGURE 2. The boxplot of case (i); the target optimal values are described by dotted lines.

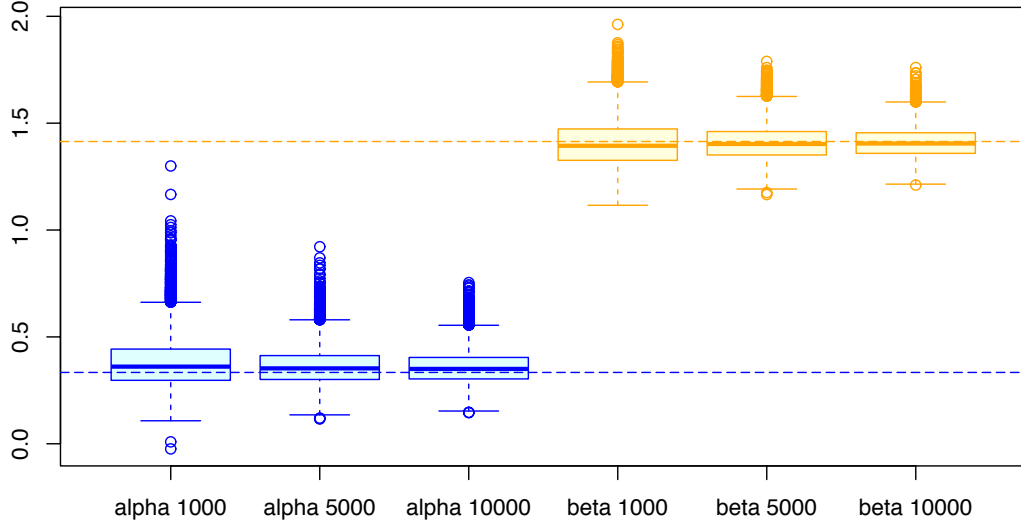
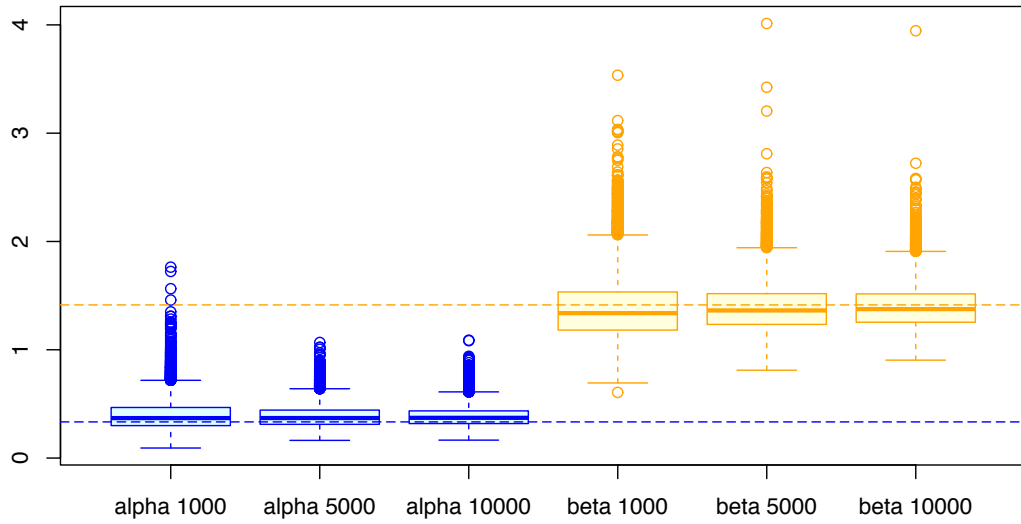


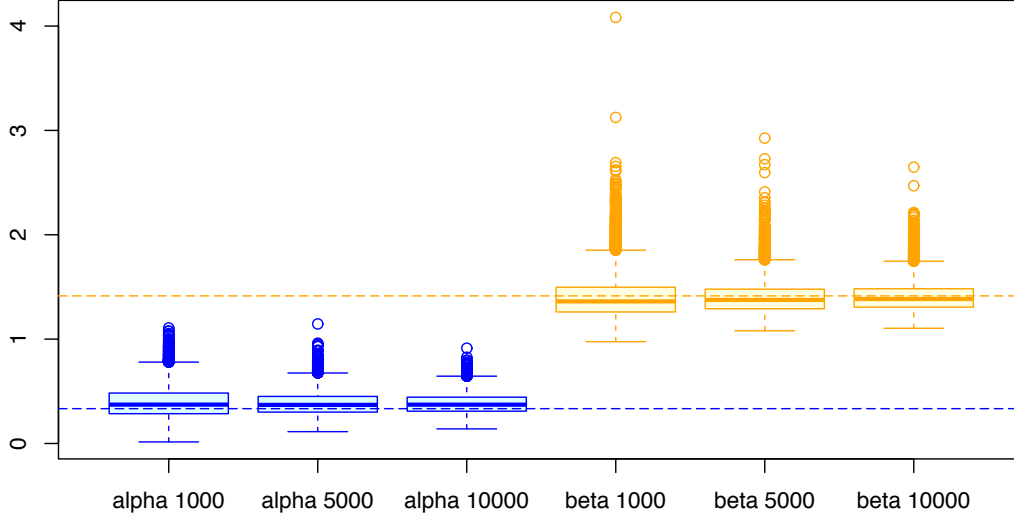
FIGURE 3. The boxplot of case (ii); the target optimal values are described by dotted lines.



mean-variance mixture of inverse Gaussian random variable, and *bGamma* (bilateral Gamma) random variable is defined by the difference of two independent Gamma random variables. For their technical accounts, we refer to [5] and [18]. To visually observe their non-Gaussianity, each density function at $t = 1$ is plotted with the density of $N(0, 1)$ in Figure 1 altogether. By taking the limit of (3.6), the characteristic function $\hat{p}(\cdot)$ of the invariant measure π_0 is given by

$$(4.1) \quad \hat{p}(u) = \exp \left\{ \int_0^\infty \psi \left(\exp \left(-\frac{s}{2} \right) u \right) ds \right\},$$

FIGURE 4. The boxplot of case (iii); the target optimal values are described by dotted lines.



where $\psi(u) := \log E[\exp(iuZ_1)]$. Differentiating \hat{p} , we have $\tilde{\kappa}_j = 2\kappa_j/j$ for the j -th cumulant $\tilde{\kappa}_j$ (resp. κ_j) of $Y \sim \pi_0$ (resp. Z_1). Hence we obtain

$$\mathbb{G}_1(\gamma) = -2 \log \gamma - \frac{2}{\gamma^2} + \int_{\mathbb{R}} \log(1+x^2) \pi_0(dx),$$

$$\mathbb{G}_2(\alpha) = -\frac{1}{\gamma^*} \left\{ \frac{1}{4} \int_{\mathbb{R}} x^3 \pi_0(dx) + \alpha \left(1 - \int_{\mathbb{R}} x^3 \pi_0(dx) + \int_{\mathbb{R}} x^4 \pi_0(dx) \right) + \alpha^2 \left(3 - 2 \int_{\mathbb{R}} x^3 \pi_0(dx) + \int_{\mathbb{R}} x^4 \pi_0(dx) \right) \right\}.$$

By solving the estimating equations, the target optimal values are given by

$$\gamma^* = \sqrt{2}, \quad \alpha^* = \frac{1 - \int_{\mathbb{R}} x^3 \pi_0(dx) + \int_{\mathbb{R}} x^4 \pi_0(dx)}{2(3 - 2 \int_{\mathbb{R}} x^3 \pi_0(dx) + \int_{\mathbb{R}} x^4 \pi_0(dx))}.$$

In the calculation, we used $\int_{\mathbb{R}} x \pi_0(dx) = 0$ and $\int_{\mathbb{R}} x^2 \pi_0(dx) = 1$. Thus, in each case, the optimal value $\theta^* := (\alpha^*, \gamma^*)$ is given as follows: (i) $\theta^* = (803/2406, \sqrt{2}) \approx (0.3337, 1.4142)$, (ii) $\theta^* = (11/30, \sqrt{2}) \approx (0.3667, 1.4142)$, and (iii) $\theta^* = (609/1658, \sqrt{2}) \approx (0.3673, 1.4142)$ (we write approximated values obtained by rounding off θ^* to four decimal places). Solving the corresponding estimating equations, our staged GQMLE are calculated as:

$$\hat{\alpha}_n = -\frac{\sum_{j=1}^n (X_{j-1} - 1)(X_j - X_{j-1})(X_{j-1}^2 + 1)}{h_n \sum_{j=1}^n (X_j - 1)^2 (X_{j-1}^2 + 1)}, \quad \hat{\gamma}_n = \sqrt{\frac{1}{nh_n} \sum_{j=1}^n (X_j - X_{j-1})^2 (X_{j-1}^2 + 1)}.$$

We generated 10000 paths of each SDE based on Euler-Maruyama scheme and constructed the estimators along with the above expressions, independently. In generating the small time increments of the driving noises, we used the function `rng` equipped to YUIMA package in R [7]. Together with the diffusion case ($\theta^* = (1/3, \sqrt{2}) \approx (0.3333, 1.4142)$), the mean and standard deviation of each estimator is shown in Table 2 where n and $h_n = 5n^{-2/3}$ denote the sample size and observation interval, respectively. We also present their boxplots to enhance the visibility. We can observe the followings from the table and boxplots:

- Overall, the estimation accuracy of $\hat{\theta}_n$ improves as T_n and n increase and h_n decrease, and this tendency reflects our main result.
- The result of case (i) is almost the same as the diffusion case. This is thought to be based on the well-known fact that $NIG(\delta, 0, \delta t, 0)$ tends to $N(0, t)$ in total variation norm as $\delta \rightarrow \infty$ for

any $t > 0$. Indeed, Figure 1 shows that the density functions of $NIG(10, 0, 10, 0)$ and $N(0, 1)$ are virtually the same.

- Concerning case (ii), the standard deviation of $\hat{\gamma}_n$ is relatively worse than the other cases. This is natural because the asymptotic variance of $\hat{\gamma}_n$ includes the forth-order-moment of Z , and $bGamma(1, \sqrt{2}, 1, \sqrt{2})$ has the highest kurtosis value as can be seen from Figure 1.
- In case (iii), the performance of $\hat{\alpha}_n$ is the worst in this experiment. This may cause from the fact that only $NIG(25/3, 20/3, 9/5, -12/5)$ is not symmetric.

5. APPENDIX

Throughout of the proofs, for functions f on $\mathbb{R} \times \Theta$, we will sometimes write f_s and f_{j-1} instead of $f_s(\theta^*)$ and $f_{j-1}(\theta^*)$ just for simplicity.

Proof of Theorem 3.1 In light of our situation, it is sufficient to check the conditions [A1''], [A4'] and [A6] in [42] for $\mathbb{G}_{1,n}$ and $\mathbb{G}_{2,n}$, respectively. For the sake of convenience, we simply write $\mathbb{Y}_{1,n}(\gamma) := \mathbb{G}_{1,n}(\gamma) - \mathbb{G}_{1,n}(\gamma^*)$ and $\mathbb{Y}_{2,n}(\alpha) := \mathbb{G}_{2,n}(\alpha) - \mathbb{G}_{2,n}(\alpha^*)$ below. Without loss of generality, we can assume $p_\gamma = p_\alpha = 1$. First we treat $\mathbb{G}_{1,n}(\cdot)$. The conditions hold if we show

$$(5.1) \quad \sup_{n \in \mathbb{N}} E \left[|\sqrt{T_n} \partial_\gamma \mathbb{G}_{1,n}(\gamma^*)|^K \right] < \infty,$$

$$(5.2) \quad \sup_{n \in \mathbb{N}} E \left[|\sqrt{T_n} (\partial_\gamma^2 \mathbb{G}_{1,n}(\gamma^*) + \Gamma_\gamma)|^K \right] < \infty,$$

$$(5.3) \quad \sup_{n \in \mathbb{N}} E \left[\sup_{\gamma \in \Theta_\gamma} |\partial_\gamma^3 \mathbb{G}_{1,n}(\gamma)|^K \right] < \infty,$$

$$(5.4) \quad \sup_{n \in \mathbb{N}} E \left[\sup_{\gamma \in \Theta_\gamma} |\sqrt{T_n} (\mathbb{Y}_{1,n}(\gamma) - \mathbb{Y}_1(\gamma))|^K \right] < \infty,$$

for any $K > 0$. The first two derivatives of $\mathbb{G}_{1,n}$ are given by

$$\begin{aligned} \partial_\gamma \mathbb{G}_{1,n}(\gamma) &= -\frac{2}{T_n} \sum_{j=1}^n \left\{ \frac{\partial_\gamma c_{j-1}(\gamma)}{c_{j-1}(\gamma)} h_n - \frac{\partial_\gamma c_{j-1}(\gamma)}{c_{j-1}^3(\gamma)} (\Delta_j X)^2 \right\}, \\ \partial_\gamma^2 \mathbb{G}_{1,n}(\gamma) &= -\frac{2}{T_n} \sum_{j=1}^n \left\{ \frac{\partial_\gamma^2 c_{j-1}(\gamma) c_{j-1}(\gamma) - (\partial_\gamma c_{j-1})^2}{c_{j-1}^2(\gamma)} h_n - \frac{\partial_\gamma^2 c_{j-1}(\gamma) c_{j-1}(\gamma) - 3(\partial_\gamma c_{j-1}(\gamma))^2}{c_{j-1}^4(\gamma)} (\Delta_j X)^2 \right\}. \end{aligned}$$

We further decompose $\partial_\gamma \mathbb{G}_{1,n}(\gamma^*)$ as

$$\partial_\gamma \mathbb{G}_{1,n}(\gamma^*) = -\frac{2}{n} \sum_{j=1}^n \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} (c_{j-1}^2 - C_{j-1}^2) + \frac{2}{T_n} \sum_{j=1}^n \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} \{(\Delta_j X)^2 - h_n C_{j-1}^2\}.$$

Since the optimal value θ^* is in the interior of Θ , the interchange of the derivative and the integral implies that $\partial_\gamma c(x, \gamma^*)(c^2(x, \gamma^*) - C^2(x))/c^3(x, \gamma^*)$ is centered in the sense that its integral with respect to π_0 is 0. Thus [24, Lemma 4.3] and [27, Lemma 5.3] lead to (5.1) and (5.4). We also have

$$\begin{aligned} \partial_\gamma^2 \mathbb{G}_{1,n}(\gamma) &= -\frac{2}{n} \sum_{j=1}^n \left\{ \frac{\partial_\gamma^2 c_{j-1}(\gamma) c_{j-1}(\gamma) - (\partial_\gamma c_{j-1})^2}{c_{j-1}^2(\gamma)} - \frac{\partial_\gamma^2 c_{j-1}(\gamma) c_{j-1}(\gamma) - 3(\partial_\gamma c_{j-1}(\gamma))^2}{c_{j-1}^4(\gamma)} C_{j-1}^2 \right\} \\ &\quad + \frac{2}{T_n} \sum_{j=1}^n \frac{\partial_\gamma^2 c_{j-1}(\gamma) c_{j-1}(\gamma) - 3(\partial_\gamma c_{j-1}(\gamma))^2}{c_{j-1}^4(\gamma)} \{(\Delta_j X)^2 - h_n C_{j-1}^2\}. \end{aligned}$$

Again applying [24, Lemma 4.3] and [27, Lemma 5.3], we obtain (5.2). Via simple calculation, the third and fourth-order derivatives of $\mathbb{G}_{1,n}$ can be represented as

$$\partial_\gamma^i \mathbb{G}_{1,n}(\gamma) = \frac{1}{n} \sum_{j=1}^n g_{j-1}^i(\gamma) + \frac{1}{T_n} \sum_{j=1}^n \tilde{g}_{j-1}^i(\gamma) \{(\Delta_j X)^2 - h_n C_{j-1}^2\}, \quad \text{for } i \in \{3, 4\},$$

with the matrix-valued functions $g^i(\cdot, \cdot)$ and $\tilde{g}^i(\cdot, \cdot)$ defined on $\mathbb{R} \times \Theta_\gamma$, and these are of at polynomial growth with respect to $x \in \mathbb{R}$ uniformly in γ . Hence (5.3) follows from Sobolev's inequality (cf. [1, Theorem 1.4.2]). Thus [42, Theorem 3-(c)] leads to the tail probability estimates of $\hat{\gamma}_n$. From Taylor's expansion, we get

$$\mathbb{Y}_{2,n}(\alpha)$$

$$\begin{aligned}
&= \frac{1}{T_n} \sum_{j=1}^n \frac{2\Delta_j X(a_{j-1}(\alpha) - a_{j-1}(\alpha^*)) + h_n(a_{j-1}^2(\alpha^*) - a_{j-1}^2(\alpha))}{c_{j-1}^2(\gamma^*)} \\
&+ \left(\int_0^1 \frac{1}{(T_n)^{3/2}} \sum_{j=1}^n \{2\Delta_j X(a_{j-1}(\alpha) - a_{j-1}(\alpha^*)) + h_n(a_{j-1}^2(\alpha^*) - a_{j-1}^2(\alpha))\} \partial_\gamma c_{j-1}^{-2}(\gamma^* + u(\hat{\gamma}_n - \gamma^*)) du \right) \\
&\quad (\sqrt{T_n}(\hat{\gamma}_n - \gamma^*)) \\
&:= \tilde{\mathbb{Y}}_{2,n}(\alpha) + \mathbb{Y}_{2,n}(\alpha)(\sqrt{T_n}(\hat{\gamma}_n - \gamma^*)).
\end{aligned}$$

Sobolev's inequality leads to

$$\begin{aligned}
&E \left[\left| \sqrt{T_n} \tilde{\mathbb{Y}}_{2,n}(\alpha) \right|^K \right] \\
&\leq E \left[\sup_{\gamma \in \Theta_\gamma} \left| \frac{1}{T_n} \sum_{j=1}^n \{2\Delta_j X(a_{j-1}(\alpha) - a_{j-1}(\alpha^*)) + h_n(a_{j-1}^2(\alpha^*) - a_{j-1}^2(\alpha))\} \partial_\gamma c_{j-1}^{-2}(\gamma) \right|^K \right] \\
&\lesssim \sup_{\gamma \in \Theta_\gamma} \left\{ E \left[\left| \frac{1}{T_n} \sum_{j=1}^n \{2\Delta_j X(a_{j-1}(\alpha) - a_{j-1}(\alpha^*)) + h_n(a_{j-1}^2(\alpha^*) - a_{j-1}^2(\alpha))\} \partial_\gamma c_{j-1}^{-2}(\gamma) \right|^K \right] \right. \\
&\quad \left. + E \left[\left| \frac{1}{T_n} \sum_{j=1}^n \{2\Delta_j X(a_{j-1}(\alpha) - a_{j-1}(\alpha^*)) + h_n(a_{j-1}^2(\alpha^*) - a_{j-1}^2(\alpha))\} \partial_\gamma^2 c_{j-1}^{-2}(\gamma) \right|^K \right] \right\},
\end{aligned}$$

for $K > 1$. The last two terms of the right-hand-side are finite from [27, Lemma 5.3], and the moment bounds of the three functions $\sqrt{T_n} \partial_\alpha^i \tilde{\mathbb{Y}}_{2,n}(\alpha)$ ($i \in \{1, 2, 3\}$) can analogously be obtained. Thus combined with the tail probability estimates of $\hat{\gamma}_n$ and Schwartz's inequality, it suffices to show the conditions for

$$\begin{aligned}
\tilde{\mathbb{G}}_{2,n}(\alpha) &:= -\frac{1}{T_n} \sum_{j=1}^n \frac{(\Delta_j X - h_n a_{j-1}(\alpha))^2}{h_n c_{j-1}^2(\gamma^*)}, \\
\tilde{\mathbb{Y}}_{2,n}(\alpha) &:= \frac{1}{T_n} \sum_{j=1}^n \frac{2\Delta_j X(a_{j-1}(\alpha) - a_{j-1}(\alpha^*)) + h_n(a_{j-1}^2(\alpha^*) - a_{j-1}^2(\alpha))}{c_{j-1}^2(\gamma^*)},
\end{aligned}$$

instead of $\mathbb{G}_{2,n}(\alpha)$ and $\mathbb{Y}_{2,n}(\alpha)$, respectively. Since their estimates can be proved in a similar way to the first half, we omit the details. \square

To derive Proposition 3.4, we prepare the next lemma. For L_1 metric $d(\cdot, \cdot)$ on \mathbb{R} , we define the coupling distance $W(\cdot, \cdot)$ between any two probability measures P and Q by

$$W(P, Q) := \inf \left\{ \int_{\mathbb{R}^2} d(x, y) d\mu(x, y) : \mu \in M(P, Q) \right\} = \inf \left\{ \int_{\mathbb{R}^2} |x - y| d\mu(x, y) : \mu \in M(P, Q) \right\},$$

where $M(P, Q)$ denotes the set of all probability measures on \mathbb{R}^2 with marginals P and Q . $W(\cdot, \cdot)$ is called the probabilistic Kantorovich-Rubinstein metric (or the first Wasserstein metric). The following assertion gives the exponential estimates of $W(P_t(\cdot, \cdot), \pi_0)$:

Lemma 5.1. *If Assumption 2.5 holds, then for any $q > 1$, there exists a positive constant C_q such that for all $x \in \mathbb{R}$,*

$$W(P_t(x, \cdot), \pi_0) \leq C_q \exp(-at)(1 + |x|^q).$$

Proof. We introduce the following Lipschitz semi-norm for a suitable real-valued function f on \mathbb{R} :

$$\|f\|_L := \sup\{|f(x) - f(y)|/|x - y| : x \neq y \text{ in } \mathbb{R}\}.$$

From Kantorovich-Rubinstein theorem (cf. [9, Theorem 11.8.2]) and Assumption 2.5, it follows that for all $x \in \mathbb{R}$,

$$\begin{aligned}
W(P_t(x, \cdot), \pi_0) &= \sup \left\{ \left| \int_{\mathbb{R}} f(y) \{P_t(x, dy) - \pi_0(dy)\} \right| : \|f\|_L \leq 1 \right\} \\
&= \sup \left\{ \left| \int_{\mathbb{R}} (f(y) - f(0)) \{P_t(x, dy) - \pi_0(dy)\} \right| : \|f\|_L \leq 1 \right\}
\end{aligned}$$

$$\begin{aligned} &\leq \sup \left\{ \left| \int_{\mathbb{R}} h(y) \{P_t(x, dy) - \pi_0(dy)\} \right| : |h(y)| \leq 1 + |y|^q \right\} \\ &\leq C_q \exp(-at)(1 + |x|^q). \end{aligned}$$

□

Proof of Proposition 3.4 It is enough to check the conditions of [38, Theorem 3.1.1 and Theorem 3.1.3] for $p_\gamma = p_\alpha = 1$. As was mentioned in the proof of Theorem 3.1, $g_1(x) := -\partial_\gamma c(x, \gamma^*)(c^2(x, \gamma^*) - C^2(x))/c^3(x, \gamma^*)$ and $g_2(x) := -\partial_\alpha a(x, \alpha^*)(A(x) - a(x, \alpha^*))/c^2(x, \gamma^*)$ are centered. In the following, we give the proof concerning g_1 and omit its index 1 for simplicity. The regularity conditions on the coefficients imply that there exist positive constants L and D such that $|g(x) - g(y)| \leq D(2 + |x|^L + |y|^L)|x - y|$. Making use of the trivial inequalities $|x - y|^l \leq |x|^l + |y|^l$ and $|x|^l \leq (1 \vee |x|^{L+l})$ for any $l \in (0, 1)$ and $x, y \in \mathbb{R}$, we have

$$\sup_{x \neq y} \frac{|g(x) - g(y)|}{(2 + |x|^{L+1-1/p} + |y|^{L+1-1/p})|x - y|^{1/p}} < \infty, \quad \text{for any } p > 1.$$

Recall that we put $h_L(x) = 1 + |x|^L$ in Assumption 2.5. The inequality (2.4) gives

$$\int_{\mathbb{R}} h_L(y) P_t(x, dy) \leq \|P_t(x, \cdot) - \pi(\cdot)\|_{h_L} + \int_{\mathbb{R}} (1 + |y|^L) \pi_0(dy) \leq \left(C_L + \int_{\mathbb{R}} (1 + |y|^L) \pi_0(dy) \right) h_L(x).$$

We write $L' = L + 1 - 1/p$ for abbreviation. Building on this estimate and the previous lemma, the conditions of [38, Theorem 3.1.1 and Theorem 3.1.3] are satisfied with

$$\begin{aligned} p &= p, q = \frac{p}{p-1}, d(x, y) = |x - y|, r(t) = \exp(-at), \phi(x) = 1 + |x|^{L'}, \\ \psi(x) &= 2^{q-1} \left(C_{qL'} + \int_{\mathbb{R}} h_{qL'}(y) \pi_0(dy) \right) h_{qL'}(x), \\ \chi(x) &= 2^{q^2-1} \left(C_{qL'} + \int_{\mathbb{R}} h_{qL'}(y) \pi_0(dy) \right)^q \left(C_{q^2L'} + \int_{\mathbb{R}} h_{q^2L'}(y) \pi_0(dy) \right) h_{q^2L'}(x), \end{aligned}$$

and here these symbols correspond to the ones used in [38]. As for g_2 , the conditions can be checked as well. Hence the desired result follows. □

To derive the asymptotic normality of $\hat{\theta}_n$, the following CLT-type theorem for stochastic integrals with respect to Poisson random measures will come into the picture:

Lemma 5.2. *Let $N(ds, dz)$ be a Poisson random measure associated with one-dimensional Lévy process defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ whose Lévy measure is written as ν_0 . Assume that a continuous vector-valued function f on $\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}$ and a \mathcal{F}_t -predictable process H_t satisfy:*

- (1) $E \left[\int_0^T \int_{\mathbb{R}} |f(T, H_s, z)|^k \nu_0(dz) ds \right] < \infty$ for all $T > 0$ and $k = 2, 4$ and there exists a nonnegative matrix C such that $E \left[\int_0^T \int_{\mathbb{R}} f(T, H_s, z)^{\otimes 2} \nu_0(dz) ds \right] \rightarrow C$ as $T \rightarrow \infty$;
- (2) there exists $\delta > 0$ such that $E \left[\int_0^T \int_{\mathbb{R}} |f(T, H_s, z)|^{2+\delta} \nu_0(dz) ds \right] \rightarrow 0$ as $T \rightarrow \infty$.

Then $\int_0^T \int_{\mathbb{R}} f(T, H_s, z) \tilde{N}(ds, dz) \xrightarrow{L} N(0, C)$ as $T \rightarrow \infty$ for the associated compensated Poisson random measure $\tilde{N}(ds, dz)$.

Proof. By Cramer-Wald device, it is sufficient to show only one-dimensional case. This proof is almost the same as [8, Theorem 14. 5. I]. For notational brevity, we set

$$X_1(t) := \int_0^t \int_{\mathbb{R}} f(T, H_s, z) \tilde{N}(ds, dz), \quad X_2(t) := \int_0^t \int_{\mathbb{R}} |f(T, H_s, z)|^2 \nu_0(dz) ds,$$

Introduce a stopping time $S := \inf\{t > 0 : X_2(t) \geq C\}$. Note that $X_2(S) = C$ because $X_2(t)$ is continuous. Define a random function $\zeta(u, t)$ by

$$\zeta(u, t) = \exp \left\{ iuX_1(t \wedge S) + \frac{u^2}{2} X_2(t \wedge S) \right\}.$$

Applying Itô's formula, we obtain

$$\zeta(u, T) = 1 + iu \int_0^{T \wedge S} \zeta(u, s-) dX_1(s) + \frac{u^2}{2} \int_0^{T \wedge S} \zeta(u, s-) dX_2(s)$$

$$\begin{aligned}
& + \sum_{0 < s \leq T \wedge S} (\zeta(u, s-) \exp \{iu \Delta X_1(s)\} - \zeta(u, s-) - iu \zeta(u, s-) \Delta X_1(s)) \\
& = 1 + \int_0^{T \wedge S} \int_{\mathbb{R}} \zeta(u, s-) (\exp \{iuf(T, H_s, z)\} - 1) \tilde{N}(ds, dz) \\
& + \int_0^{T \wedge S} \int_{\mathbb{R}} \zeta(u, s-) \left(\exp \{iuf(T, H_s, z)\} - 1 - iuf(T, H_s, z) + \frac{u^2}{2} |f(T, H_s, z)|^2 \right) \nu_0(dz) ds.
\end{aligned}$$

For later use, we here present the following elementary inequality (cf. [10]): for all $u \in \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$,

$$(5.5) \quad \left| \exp(iu) - \sum_{j=0}^n \frac{(iu)^j}{j!} \right| \leq \frac{|u|^{n+1}}{(n+1)!} \wedge \frac{2|u|^n}{n!}.$$

By the definition of S , we have $|\zeta(u, T)| \leq \exp \{u^2 C/2\}$. Since $\int_0^T \int_{\mathbb{R}} \zeta(u, s-) (\exp \{iuf(T, H_s, z)\} - 1) \tilde{N}(ds, dz)$ is an L_2 -martingale (cf. [2, Section 4]) from these estimates, the optional sampling theorem implies that

$$E \left[\int_0^{T \wedge S} \int_{\mathbb{R}} \zeta(u, s-) (\exp \{iuf(T, H_s, z)\} - 1) \tilde{N}(ds, dz) \right] = 0.$$

Next we show that

$$E \left[\int_0^{T \wedge S} \int_{\mathbb{R}} \zeta(u, s-) \left(\exp \{iuf(T, H_s, z)\} - 1 - iuf(T, H_s, z) + \frac{u^2}{2} |f(T, H_s, z)|^2 \right) \nu_0(dz) ds \right] \rightarrow 0.$$

Again using the above estimates, we have

$$\begin{aligned}
& \left| E \left[\int_0^{T \wedge S} \int_{\mathbb{R}} \zeta(u, s-) \left(\exp \{iuf(T, H_s, z)\} - 1 - iuf(T, H_s, z) + \frac{u^2}{2} |f(T, H_s, z)|^2 \right) \nu_0(dz) ds \right] \right| \\
& \leq E \left[\int_0^{T \wedge S} \int_{\mathbb{R}} \exp \left\{ \frac{u^2}{2} C \right\} \left(\frac{|uf(T, H_s, z)|^3}{6} \wedge |uf(T, H_s, z)|^2 \right) \nu_0(dz) ds \right] \\
& \leq C_\delta \exp \left\{ \frac{u^2}{2} C \right\} E \left[\int_0^T \int_{\mathbb{R}} |uf(T, H_s, z)|^{2+\delta} \nu_0(dz) ds \right] \rightarrow 0,
\end{aligned}$$

where C_δ is a positive constant such that $|x|^3/6 \wedge |x|^2 \leq C_\delta |x|^{2+\delta}$ for all $x \in \mathbb{R}$. At last we observe that $X_1(T \wedge S) - X_1(T) \xrightarrow{p} 0$. In view of Lenglart's inequality and the isometry property of stochastic integral with respect to Poisson random measure (cf. [2, Section 4]), it suffices to show $E[\int_{T \wedge S}^T \int_{\mathbb{R}} |f(T, H_s, z)|^2 \nu_0(dz) ds] \rightarrow 0$. However the latter convergence is clear from Assumption (1). Hence the proof is complete. \square

Next we show the following lemma which gives the fundamental small time moment estimate of X :

Lemma 5.3. *Under Assumptions 2.3 and 2.4, it follows that*

$$(5.6) \quad E^{j-1}[|X_s - X_{j-1}|^p] \lesssim h_n(1 + |X_{j-1}|^K),$$

for any positive constant $p \in (1 \vee \beta, 2)$ and $s \in (t_{j-1}, t_j]$.

Proof. Recall that $\int |z|^p \nu_0(dz) < \infty$ from Assumption 2.3. By Lipschitz continuity of the coefficients and [11, Theorem 1.1], it follows that

$$\begin{aligned}
& E^{j-1}[|X_s - X_{j-1}|^p] \\
& \lesssim E^{j-1} \left[\left| \int_{t_{j-1}}^s (A_s - A_{j-1}) ds + \int_{t_{j-1}}^s (C_{s-} - C_{j-1}) dZ_s \right|^p + h_n^p |A_{j-1}|^p + h_n |C_{j-1}|^p \int_{\mathbb{R}} |z|^p \nu_0(dz) \right] + o_p(h_n) \\
& \lesssim h_n(1 + |X_{j-1}|^K + o_p(1)) + h_n^{p-1} \int_{t_{j-1}}^{t_j} E^{j-1}[|X_s - X_{j-1}|^p] ds + E^{j-1} \left[\left| \int_{t_{j-1}}^{t_j} (C_{s-} - C_{j-1}) dZ_s \right|^p \right].
\end{aligned}$$

Applying Burkholder-Davis-Gundy's inequality (cf. [32, Theorem 48]), we have

$$E^{j-1} \left[\left| \int_{t_{j-1}}^{t_j} (C_{s-} - C_{j-1}) dZ_s \right|^p \right] \lesssim E^{j-1} \left[\left(\int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} (C_{s-} - C_{j-1})^2 z^2 N(ds, dz) \right)^{p/2} \right]$$

$$\begin{aligned}
&= E^{j-1} \left[\left(\sum_{t_{j-1} \leq s < t_j} (C_{s-} - C_{j-1})^2 (Z_s - Z_{s-})^2 \right)^{p/2} \right] \\
&\leq E^{j-1} \left[\sum_{t_{j-1} \leq s < t_j} |C_{s-} - C_{j-1}|^p |Z_s - Z_{s-}|^p \right] \\
&= \int_{t_{j-1}}^{t_j} E^{j-1} [|X_s - X_{j-1}|^p] ds \int_{\mathbb{R}} |z|^p \nu_0(dz),
\end{aligned}$$

for the Poisson random measure $N(ds, dz)$ associated with Z . Hence Gronwall's inequality gives (5.6). \square

Proof of Theorem 3.6 According to Cramer-Wald device, it is enough to show for $p_\gamma = p_\alpha = 1$. From a similar estimates used in Theorem 3.1, we have

$$\begin{aligned}
\sqrt{T_n} \partial_\gamma \mathbb{G}(\gamma^*) &= -\frac{2}{\sqrt{T_n}} \sum_{j=1}^n \left\{ \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} (h_n c_{j-1}^2 - (\Delta_j X)^2) \right\} \\
&= -\frac{2}{\sqrt{T_n}} \sum_{j=1}^n \left\{ \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} (h_n c_{j-1}^2 - C_{j-1}^2 (\Delta_j Z)^2) \right\} + o_p(1) \\
&= -\frac{2}{\sqrt{T_n}} \sum_{j=1}^n \left\{ \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} C_{j-1}^2 (h_n - (\Delta_j Z)^2) \right\} - \frac{2}{\sqrt{T_n}} \int_0^{T_n} \frac{\partial_\gamma c_s}{c_s^3} (c_s^2 - C_s^2) ds \\
&\quad - \frac{2}{\sqrt{T_n}} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \left\{ \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} (c_{j-1}^2 - C_{j-1}^2) - \frac{\partial_\gamma c_s}{c_s^3} (c_s^2 - C_s^2) \right\} ds + o_p(1) \\
(5.7) \quad &= \mathbb{F}_{1,n} + \mathbb{F}_{2,n} + \mathbb{F}_{3,n} + o_p(1).
\end{aligned}$$

We evaluate each term separately below. Rewriting $\mathbb{F}_{1,n}$ in a stochastic integral form via Itô's formula, we have

$$\begin{aligned}
\mathbb{F}_{1,n} &= -\frac{2}{\sqrt{T_n}} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \frac{\partial_\gamma c_{s-}}{c_{s-}^3} C_{s-}^2 z^2 \tilde{N}(ds, dz) - \frac{2}{\sqrt{T_n}} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} C_{j-1}^2 - \frac{\partial_\gamma c_{s-}}{c_{s-}^3} C_{s-}^2 \right) z^2 \tilde{N}(ds, dz) \\
&\quad - \frac{4}{\sqrt{T_n}} \sum_{j=1}^n \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} C_{j-1}^2 \int_{t_{j-1}}^{t_j} (Z_{s-} - Z_{j-1}) dZ_s.
\end{aligned}$$

for the compensated Poisson random measure $\tilde{N}(ds, dz)$ associated with Z . Using Burkholder's inequality and the isometry property, it follows that

$$\begin{aligned}
&E \left[\left(\frac{1}{\sqrt{T_n}} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} C_{j-1}^2 - \frac{\partial_\gamma c_{s-}}{c_{s-}^3} C_{s-}^2 \right) z^2 \tilde{N}(ds, dz) \right)^2 \right] \\
&\lesssim \frac{1}{T_n} \sum_{j=1}^n E \left[\left(\int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3} C_{j-1}^2 - \frac{\partial_\gamma c_{s-}}{c_{s-}^3} C_{s-}^2 \right) z^2 \tilde{N}(ds, dz) \right)^2 \right] \\
&\lesssim \frac{1}{T_n} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} E \left[\left(\int_0^1 \partial_x \left(\frac{\partial_\gamma c}{c^3} C^2 \right) (X_{j-1} + u(X_s - X_{j-1})) du \right) (X_s - X_{j-1}) \right] ds \\
&\lesssim \frac{1}{T_n} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \sqrt{\sup_{t>0} E[1 + |X_t|^K]} \sqrt{E[(X_s - X_{j-1})^2]} ds \\
&\lesssim \sqrt{h_n},
\end{aligned}$$

and that

$$E \left[\left| \int_{t_{j-1}}^{t_j} (J_{s-} - J_{j-1}) dJ_s \right|^2 \right] \lesssim \int_{t_{j-1}}^{t_j} E[|J_{s-t_{j-1}}|^2] ds \leq h_n^2.$$

Hence

$$\mathbb{F}_{1,n} = -\frac{2}{\sqrt{T_n}} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \frac{\partial_{\gamma} c_{s-}}{c_{s-}^3} C_{s-}^2 z^2 \tilde{N}(ds, dz) + o_p(1).$$

Let us turn to observe $\mathbb{F}_{2,n}$. Let $f_{i,t} := f_i(X_t)$ for $i = 1, 2$, and especially, let $f_{i,j} := f_i(X_{t_j})$. From Proposition 3.4, we obtain

$$\begin{aligned} \mathbb{F}_{2,n} &= -\frac{2}{\sqrt{T_n}} \sum_{j=1}^n \left(f_{1,j} - f_{1,j-1} + \int_{t_{j-1}}^{t_j} \frac{\partial_{\gamma} c_s}{c_s^3} (c_s^2 - C_s^2) ds \right) - \frac{2}{\sqrt{T_n}} (f_{1,n} - f_{1,0}) \\ &= -\frac{2}{\sqrt{T_n}} \sum_{j=1}^n \left(f_{1,j} - f_{1,j-1} + \int_{t_{j-1}}^{t_j} \frac{\partial_{\gamma} c_s}{c_s^3} (c_s^2 - C_s^2) ds \right) + o_p(1). \end{aligned}$$

For abbreviation, we simply write $\xi_{1,j}(t) = f_{1,t} - f_{1,j-1} + \int_{t_{j-1}}^t \partial_{\gamma} c_s (c_s^2 - C_s^2) / c_s^3 ds$ below. According to Proposition 3.4, the weighted Hölder continuity of f , and Lemma 5.3, $\{\xi_{1,j}(t), \mathcal{F}_{t_{j-1}+t} : t \in [0, h_n]\}$ turns out to be an L_2 -martingale. Thus the martingale representation theorem [15, Theorem III. 4. 34] implies that there exists a predictable process $s \mapsto \tilde{\xi}_{1,j}(s, z)$ such that

$$\xi_{1,j}(t) = \int_{t_{j-1}}^t \int_{\mathbb{R}} \tilde{\xi}_{1,j}(s, z) \tilde{N}(ds, dz).$$

Hence the continuous martingale component of $\xi_{1,j}$ is 0. By the property of f , we can define the stochastic integral $\int_{t_{j-1}}^t \int_{\mathbb{R}} (f_1(X_{s-} + C_{s-}z) - f_1(X_{s-})) \tilde{N}(ds, dz)$ on $t \in [t_{j-1}, t_j]$ and this process is also an L_2 -martingale with respect to $\{\mathcal{F}_{t_{j-1}+t} : t \in [0, h_n]\}$. Utilizing [15, Theorem I. 4. 52] and [32, Corollary II. 6. 3], we have

$$\begin{aligned} &E \left[\left| \frac{1}{\sqrt{T_n}} \sum_{j=1}^n \left\{ \xi_{1,j}(t_j) - \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} (f_1(X_{s-} + C_{s-}z) - f_1(X_{s-})) \tilde{N}(ds, dz) \right\} \right|^2 \right] \\ &\lesssim \frac{1}{T_n} \sum_{j=1}^n E \left[\left| \xi_{1,j}(t_j) - \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} (f_1(X_{s-} + C_{s-}z) - f_1(X_{s-})) \tilde{N}(ds, dz) \right|^2 \right] \\ &= \frac{1}{T_n} \sum_{j=1}^n E \left[\left[\xi_{1,j}(\cdot) - \int_{t_{j-1}}^{\cdot} \int_{\mathbb{R}} (f_1(X_{s-} + C_{s-}z) - f_1(X_{s-})) \tilde{N}(ds, dz) \right]_{t_j} \right] = 0. \end{aligned}$$

Here $[Y]_t$ denotes the quadratic variation for any semimartingale Y at time t , and we used Burkholder's inequality for a martingale difference between the first line and the second line. By similar estimates above, we have $\mathbb{F}_{3,n} = o_p(1)$. Having these arguments in hand, it turns out that

$$\sqrt{T_n} \partial_{\gamma} \mathbb{G}(\gamma^*) = -\frac{2}{\sqrt{T_n}} \int_0^{T_n} \int_{\mathbb{R}} \left(\frac{\partial_{\gamma} c_{s-}}{c_{s-}^3} C_{s-}^2 z^2 + f_1(X_{s-} + C_{s-}z) - f_1(X_{s-}) \right) \tilde{N}(ds, dz) + o_p(1).$$

We can deduce from Assumption 2.4 and Proposition 3.4 that there exist positive constants K, K', K'' and $\epsilon_0 < 1 \wedge (2 - \beta)$ such that for all $z \in \mathbb{R}$

$$\begin{aligned} &\sup_t \left\{ \frac{1}{t} \int_0^t E \left[\left(\frac{\partial_{\gamma} c_s}{c_s^3} C_s^2 z^2 + f_1(X_s + C_s z) - f_1(X_s) \right)^2 \right] ds \right\} \\ &\lesssim \sup_t \left\{ \frac{1}{t} \int_0^t (|z|^{2-\epsilon_0} \vee z^4) (1 + \sup_t E[|X_t|^K] + (1 + \sup_t E[|X_t|^{K'}]) |z|^{K''}) ds \right\} \\ &\lesssim (|z|^{2-\epsilon_0} \vee z^4) (1 + |z|^{K''}), \end{aligned}$$

and the last term is ν_0 -integrable. Then, there exist positive constants K and K' (possibly take different values from the previous ones) such that for any $z \in \mathbb{R}$,

$$\begin{aligned} &\left| \frac{1}{t} \int_0^t E \left[\left(\frac{\partial_{\gamma} c_s}{c_s^3} C_s^2 z^2 + f_1(X_s + C_s z) - f_1(X_s) \right)^2 \right] ds \right. \\ &\quad \left. - \int_{\mathbb{R}} \left(\frac{\partial_{\gamma} c(y, \gamma^*)}{c^3(y, \gamma^*)} C^2(y) z^2 + f_1(y + C(y)z) - f_1(y) \right)^2 \pi_0(dy) \right| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{t} \int_0^t \int_{\mathbb{R}} \int_{\mathbb{R}} \left| \left(\frac{\partial_{\gamma} c(y, \gamma^*)}{c^3(y, \gamma^*)} C^2(y) z^2 + f_1(y + C(y)z) - f_1(y) \right)^2 \right| (P_t(x, dy) - \pi_0(dy)) \eta(dx) ds \\
&\lesssim (|z|^{2-\epsilon_0} \vee z^4) (1 + |z|^{K'}) \frac{1}{t} \int_0^t \int_{\mathbb{R}} \|P_t(x, \cdot) - \pi_0(\cdot)\|_{h_K} \eta(dx) ds \\
&\rightarrow 0.
\end{aligned}$$

Thus the dominated convergence theorem and the isometry property give

$$\begin{aligned}
&\lim_{n \rightarrow \infty} E \left[\left(\frac{1}{\sqrt{T_n}} \int_0^{T_n} \int_{\mathbb{R}} \left(\frac{\partial_{\gamma} c_{s-}}{c_{s-}^3} C_{s-}^2 z^2 + f_1(X_{s-} + C_{s-}z) - f_1(X_{s-}) \right) \tilde{N}(ds, dz) \right)^2 \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{T_n} \int_0^{T_n} \int_{\mathbb{R}} E \left[\left(\frac{\partial_{\gamma} c_{s-}}{c_{s-}^3} C_{s-}^2 z^2 + f_1(X_{s-} + C_{s-}z) - f_1(X_{s-}) \right)^2 \right] \nu_0(dz) ds \\
&= \frac{1}{4} \Sigma_{\gamma}.
\end{aligned}$$

It follows from Assumption 2.5 and Proposition 3.4 that

$$\lim_{n \rightarrow \infty} E \left[\int_0^{T_n} \int_{\mathbb{R}} \left\{ \frac{1}{\sqrt{T_n}} \left(\frac{\partial_{\gamma} c_{s-}}{c_{s-}^3} C_{s-}^2 z^2 + f_1(X_{s-} + C_{s-}z) - f_1(X_{s-}) \right) \right\}^{2+K} \nu_0(dz) ds \right] \rightarrow 0.$$

From Taylor expansion around γ^* , $\partial_{\alpha} \mathbb{G}_{2,n}(\alpha)$ is decomposed as:

$$\begin{aligned}
\sqrt{T_n} \partial_{\alpha} \mathbb{G}_{2,n}(\alpha^*) &= \frac{2}{\sqrt{T_n}} \sum_{j=1}^n \frac{\partial_{\alpha} a_{j-1}}{c_{j-1}^2} (\Delta_j X - h_n a_{j-1}) + \frac{2}{T_n} \sum_{j=1}^n \partial_{\alpha} a_{j-1} (\Delta_j X - h_n a_{j-1}) \partial_{\gamma} c_{j-1}^{-2} (\sqrt{T_n}(\hat{\gamma}_n - \gamma^*)) \\
&\quad + \left(\int_0^1 \frac{2}{(T_n)^{3/2}} \sum_{j=1}^n \partial_{\alpha} a_{j-1} (\Delta_j X - h_n a_{j-1}) \partial_{\gamma}^2 c_{j-1}^{-2} (\gamma^* + u(\hat{\gamma}_n - \gamma^*)) du \right) (\sqrt{T_n}(\hat{\gamma}_n - \gamma^*))^2
\end{aligned}$$

Sobolev's inequality and the tail probability estimates of $\hat{\gamma}_n$ imply that the third term of the right-hand-side is $o_p(1)$. Hence a similar manner to the first half leads to

$$\begin{aligned}
&\sqrt{T_n} \partial_{\alpha} \mathbb{G}_{2,n}(\alpha^*) - \frac{2}{T_n} \sum_{j=1}^n \partial_{\alpha} a_{j-1} (\Delta_j X - h_n a_{j-1}) \partial_{\gamma} c_{j-1}^{-2} (\sqrt{T_n}(\hat{\gamma}_n - \gamma^*)) \\
&= \frac{2}{\sqrt{T_n}} \sum_{j=1}^n \frac{\partial_{\alpha} a_{j-1}}{c_{j-1}^2} (\Delta_j X - h_n a_{j-1}) + o_p(1) \\
&= \frac{2}{\sqrt{T_n}} \sum_{j=1}^n \frac{\partial_{\alpha} a_{j-1}}{c_{j-1}^2} C_{j-1} \Delta_j Z + \frac{2}{\sqrt{T_n}} \int_0^{T_n} \frac{\partial_{\alpha} a_s}{c_s^2} (A_s - a_s) ds + o_p(1) \\
&= \frac{2}{\sqrt{T_n}} \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \frac{\partial_{\alpha} a_s}{c_{s-}^2} C_{s-} \tilde{N}(ds, dz) + \frac{2}{\sqrt{T_n}} \sum_{j=1}^n \left(f_{2,j} - f_{2,j-1} + \int_{t_{j-1}}^{t_j} \frac{\partial_{\alpha} a_s}{c_s^2} (A_s - a_s) ds \right) + o_p(1) \\
&= \frac{2}{\sqrt{T_n}} \int_0^{T_n} \int_{\mathbb{R}} \left(\frac{\partial_{\alpha} a_s}{c_{s-}^2} C_{s-} z + f_2(X_{s-} + C_{s-}z) - f_2(X_{s-}) \right) \tilde{N}(ds, dz) + o_p(1),
\end{aligned}$$

and we have

$$\begin{aligned}
&\lim_{n \rightarrow \infty} E \left[\left(\frac{1}{\sqrt{T_n}} \int_0^{T_n} \int_{\mathbb{R}} \left(\frac{\partial_{\alpha} a_s}{c_{s-}^2} C_{s-} z + f_2(X_{s-} + C_{s-}z) - f_2(X_{s-}) \right) \tilde{N}(ds, dz) \right)^2 \right] = \frac{1}{4} \Sigma_{\alpha}, \\
&\lim_{n \rightarrow \infty} E \left[\int_0^{T_n} \int_{\mathbb{R}} \left\{ \frac{1}{\sqrt{T_n}} \left(\frac{\partial_{\alpha} a_s}{c_{s-}^2} C_{s-} z + f_2(X_{s-} + C_{s-}z) - f_2(X_{s-}) \right) \right\}^{2+K} \nu_0(dz) ds \right] = 0.
\end{aligned}$$

From the isometry property and the trivial identity $xy = \{(x+y)^2 - (x-y)^2\}/4$ for any $x, y \in \mathbb{R}$, it follows that

$$\lim_{n \rightarrow \infty} E \left[\left(\frac{1}{\sqrt{T_n}} \int_0^{T_n} \int_{\mathbb{R}} \left(\frac{\partial_{\gamma} c_{s-}}{c_{s-}^3} C_{s-}^2 z^2 + f_1(X_{s-} + C_{s-}z) - f_1(X_{s-}) \right) \tilde{N}(ds, dz) \right)^2 \right]$$

$$\times \left(\frac{1}{\sqrt{T_n}} \int_0^{T_n} \int_{\mathbb{R}} \left(\frac{\partial_{\alpha} a_s}{c_{s-}^2} C_{s-} z + f_2(X_{s-} + C_{s-} z) - f_2(X_{s-}) \right) \tilde{N}(ds, dz) \right) \Bigg] = -\frac{1}{4} \Sigma_{\alpha\gamma}.$$

Hence the moment estimates in the proof of Theorem 3.1, Lemma 5.2 and Taylor's formula yield that

$$\begin{aligned} & \sqrt{T_n} \begin{pmatrix} -\partial_{\gamma}^2 \mathbb{G}_{1,n}(\gamma^*) & 0 \\ -\frac{2}{T_n} \sum_{j=1}^n \partial_{\alpha} a_{j-1} (\Delta_j X - h_n a_{j-1}) \partial_{\gamma} c_{j-1}^{-2} & -\partial_{\alpha}^2 \mathbb{G}_{2,n}(\alpha^*) \end{pmatrix} \begin{pmatrix} \hat{\gamma}_n - \gamma^* \\ \hat{\alpha}_n - \alpha^* \end{pmatrix} \\ &= \sqrt{T_n} \begin{pmatrix} \partial_{\gamma} \mathbb{G}_{1,n}(\gamma^*) \\ \partial_{\alpha} \mathbb{G}_{2,n}(\alpha^*) \end{pmatrix} + o_p(1) \xrightarrow{\mathcal{L}} N(0, \Sigma). \end{aligned}$$

To achieve the desired result, it suffices to show

$$-\partial_{\gamma}^2 \mathbb{G}_{1,n}(\gamma^*) \xrightarrow{P} \Gamma_{\gamma}, \quad -\partial_{\alpha}^2 \mathbb{G}_{2,n}(\alpha^*) \xrightarrow{P} \Gamma_{\alpha}, \quad \text{and} \quad -\frac{2}{T_n} \sum_{j=1}^n \partial_{\alpha} a_{j-1} (\Delta_j X - h_n a_{j-1}) \partial_{\gamma} c_{j-1}^{-2} \xrightarrow{P} \Gamma_{\alpha\gamma}.$$

However the first two convergence are straightforward from the proof of Theorem 3.1, and the last convergence follows from ergodic theorem. Thus the proof is complete.

Acknowledgement. The author thanks to Professor H. Masuda for his constructive comments. He is also grateful to the valuable advice of Professor A. M. Kulik about extended Poisson equation. This work was supported by JST CREST Grant Number JPMJCR14D7, Japan.

REFERENCES

- [1] R. A. Adams. Some integral inequalities with applications to the imbedding of Sobolev spaces defined over irregular domains. *Trans. Amer. Math. Soc.*, 178:401–429, 1973.
- [2] D. Applebaum. *Lévy processes and stochastic calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2009.
- [3] G. Barles, R. Buckdahn, and E. Pardoux. Backward stochastic differential equations and integral-partial differential equations. *Stochastics Stochastics Rep.*, 60(1-2):57–83, 1997.
- [4] G. Barles and C. Imbert. Second-order elliptic integro-differential equations: viscosity solutions' theory revisited. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 25(3):567–585, 2008.
- [5] O. E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance Stoch.*, 2(1):41–68, 1998.
- [6] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete*, 60(2):185–201, 1982.
- [7] A. Brouste, M. Fukasawa, H. Hino, S. M. Iacus, K. Kamatani, Y. Koike, H. Masuda, R. Nomura, T. Ogihara, Y. Shimizu, et al. Parameter estimation for the discretely observed fractional Ornstein-Uhlenbeck process and the Yuima R package. *Comput. Statist.*, 28(4):1529–1547, 2013.
- [8] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II. Probability and its Applications* (New York). Springer, New York, second edition, 2008. General theory and structure.
- [9] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [10] R. Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.
- [11] J. E. Figueroa-López. Small-time moment asymptotics for lévy processes. *Statist. Probab. Lett.*, 78(18):3355–3365, 2008.
- [12] J. E. Figueroa-López. Nonparametric estimation of Lévy models based on discrete-sampling. In *Optimality*, volume 57 of *IMS Lecture Notes Monogr. Ser.*, pages 117–146. Inst. Math. Statist., Beachwood, OH, 2009.
- [13] S. Hamadène. Viscosity solutions of second order integral-partial differential equations without monotonicity condition: A new result. *Nonlinear Anal.*, 147:213–235, 2016.
- [14] S. Hamadène and M.-A. Morlais. Viscosity solutions for second order integro-differential equations without monotonicity condition: the probabilistic approach. *Stochastics*, 88(4):632–649, 2016.
- [15] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.
- [16] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, 24(2):211–229, 1997.
- [17] T. Komorowski and A. Walczuk. Central limit theorem for Markov processes with spectral gap in the Wasserstein metric. *Stochastic Process. Appl.*, 122(5):2155–2184, 2012.
- [18] U. Küchler and S. Tappe. Bilateral gamma distributions and processes in financial mathematics. *Stochastic Process. Appl.*, 118(2):261–283, 2008.
- [19] H. Long, C. Ma, and Y. Shimizu. Least squares estimators for stochastic differential equations driven by small lévy noises. *Stochastic Processes and their Applications*, 2016.
- [20] H. Mai. Efficient maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 20(2):919–957, 2014.
- [21] H. Masuda. Ergodicity and exponential β -mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.*, 117(1):35–56, 2007.
- [22] H. Masuda. Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. *Electron. J. Stat.*, 4:525–565, 2010.
- [23] H. Masuda. Asymptotics for functionals of self-normalized residuals of discretely observed stochastic processes. *Stochastic Process. Appl.*, 123(7):2752–2778, 2013.

- [24] H. Masuda. Convergence of gaussian quasi-likelihood random fields for ergodic lévy driven sde observed at high frequency. *Ann. Statist.*, 41(3):1593–1641, 2013.
- [25] H. Masuda. Parametric estimation of Lévy processes. In *Lévy matters. IV*, volume 2128 of *Lecture Notes in Math.*, pages 179–286. Springer, Cham, 2015.
- [26] H. Masuda. Non-gaussian quasi-likelihood estimation of sde driven by locally stable Lévy process. *arXiv preprint arXiv:1608.06758*, 2016.
- [27] H. Masuda and Y. Uehara. Two-step estimation of ergodic lévy driven sde. *Stat. Inference Stoch. Process, to appear.*, 2016.
- [28] H. Masuda and Y. Uehara. On stepwise estimation of lévy driven stochastic differential equation (japanese). *Proc. Inst. Statist. Math., to appear.*, 2017.
- [29] R. Nickl, M. Reiß, J. Söhl, and M. Trabs. High-frequency Donsker theorems for Lévy measures. *Probab. Theory Related Fields*, 164(1-2):61–108, 2016.
- [30] T. Ogihara and N. Yoshida. Quasi-likelihood analysis for the stochastic differential equation with jumps. *Stat. Inference Stoch. Process.*, 14(3):189–229, 2011.
- [31] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001.
- [32] P. E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition, 2004. Stochastic Modelling and Applied Probability.
- [33] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [34] K.-i. Sato and M. Yamazato. Operator-self-decomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stochastic Process. Appl.*, 17(1):73–100, 1984.
- [35] Y. Shimizu and N. Yoshida. Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat. Inference Stoch. Process.*, 9(3):227–277, 2006.
- [36] M. Uchida and N. Yoshida. Estimation for misspecified ergodic diffusion processes from discrete observations. *ESAIM Probab. Stat.*, 15:270–290, 2011.
- [37] M. Uchida and N. Yoshida. Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Process. Appl.*, 122(8):2885–2924, 2012.
- [38] A. Y. Veretennikov and A. M. Kulik. The extended Poisson equation for weakly ergodic Markov processes. *Teor. Ėmovir. Mat. Stat.*, (85):22–38, 2011.
- [39] A. Y. Veretennikov and A. M. Kulik. Diffusion approximation of systems with weakly ergodic Markov perturbations. II. *Teor. Ėmovir. Mat. Stat.*, (88):1–16, 2013.
- [40] F.-Y. Wang. Derivative formula and Harnack inequality for SDEs driven by Lévy processes. *Stoch. Anal. Appl.*, 32(1):30–49, 2014.
- [41] F.-Y. Wang, L. Xu, and X. Zhang. Gradient estimates for SDEs driven by multiplicative Lévy noise. *J. Funct. Anal.*, 269(10):3195–3219, 2015.
- [42] N. Yoshida. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.*, 63(3):431–479, 2011.

(Yuma Uehara, corresponding author) GRADUATE SCHOOL OF MATHEMATICS, KYUSHU UNIVERSITY, 744 MOTOOKA, NISHI-KU, FUKUOKA 819-0395, JAPAN

E-mail address: y-uehara@math.kyushu-u.ac.jp